



Universidad de Carabobo Facultad Experimental
de Ciencias y Tecnología Departamento de
Computación



Aplicación para comparar precios de la competencia del sector retail de alimentos venezolano, con Web Scraping y Aprendizaje Automático. Un enfoque de Inteligencia de Negocio

Tutores:
Prof. Marylin Giugni
Prof. Mirella Herrera

Autores:
Jesús David Machado
José David López

Noviembre, 2025

Agenda

1 Problema

2 Objetivos

3 Marco teórico

4 Metodología y arquitectura

5 Resultados

6 Conclusiones y futuro

Problema

El reto del retailer en un mercado volátil



El retailer

- Necesidad de decisiones de precios rápidas y precisas.
- Procesos manuales lentos y costosos.
- Dependencia de información errónea, con un 35% de reportes tardíos, un 27% de datos inexactos y un 23% están desactualizados (Nuñez Cartolin, 2021).

El mercado

- El entorno económico muestra una inflación del 12,8% mensual y un 117% anual.
- Crecimiento acelerado del e-commerce.
- Competencia digital intensificada.

La inestabilidad económica en Venezuela, marcada por la constante fluctuación del bolívar, intensifica la volatilidad de precios en el sector retail local, obligando a los comerciantes a adaptarse rápidamente.

Problema

El manejo de los datos de forma manual

El proceso manual de recopilación de precios consume recursos valiosos y genera información obsoleta. Este flujo ineficiente impacta directamente la competitividad y rentabilidad.



Búsqueda manual

Revisión constante de webs competidoras de forma manual.

Consumo de tiempo

+160 horas/mes dedicadas a un proceso que la tecnología puede reducir a 8 segundos, liberando 20 días laborales al año (Nuñez Cartolin, 2021).



Reportes inexactos y pérdida de valor

La información desactualizada e inexacta provoca pérdida de oportunidades y menor competitividad.

Costo monetario

7000 USD es el costo mensual aproximado del proceso manual. (Nuñez Cartolin, 2021).



Justificación

Relevancia en un entorno volátil

El mercado venezolano, con una inflación anualizada del 117% (OVF, 2025), exige a las empresas del sector retail una agilidad extrema para ajustar precios. La dependencia de procesos manuales para monitorear a la competencia es insostenible y genera una desventaja competitiva crítica.

Ventajas competitivas

La solución propuesta genera ventajas significativas para la empresa al permitir:

- **Optimización dinámica de precios:**
Tomar decisiones estratégicas con información precisa del mercado en tiempo real.
- **Incremento en eficiencia:**
Automatizar la recolección y análisis de datos para reducir costos y liberar recursos.
- **Conocimiento del mercado:**
Identificar patrones y tendencias para centrar las estrategias en el cliente.

Contribución al departamento de computación

Este trabajo establece un precedente en la aplicación de técnicas avanzadas de IA, web scraping y BI para resolver problemas del contexto económico real venezolano. Crea una base sólida para futuras investigaciones y desarrollos dentro del departamento, impulsando el avance del conocimiento aplicado en la región.

Marco teórico

Trabajos relacionados

Khder (2021)

Proporcionó un marco de responsabilidad ética y legal. Su trabajo fue la guía para asegurar que la extracción de datos se realizará de forma sostenible, respetando normativas como robots.txt y los términos de servicio.

Cobo, Benítez et al. (2025)

Su estudio sirvió como una referencia para validar el enfoque integral (Scraping + BI). Su caso de éxito en supermercados colombianos demostró la viabilidad de combinar estas tecnologías, confirmando la dirección de nuestro proyecto.

Harsh Singh (2023) - Kaggle

Otorgó una base técnica para el modelo predictivo. Su metodología práctica de optimización de precios con Machine Learning fue la referencia directa para implementar y ajustar nuestro modelo Random Forest.

Brown et al. (2024)

Sus aportes ayudaron a complementar el marco ético con soluciones prácticas como el rate limiting, para evitar la sobrecarga de servidores y garantizar un comportamiento responsable.

D'Souza et al. (2024)

Proporciona estrategias para manejar la variabilidad de datos. Su estudio abordó el desafío técnico de integrar datos de múltiples sitios de e-commerce con estructuras variables, aportando soluciones para la homologación de productos y la robustez del pipeline.

Marco teórico

Bases teóricas

Web scraping

Técnica de extracción automatizada de datos desde sitios web para su posterior análisis en un formato estructurado.

Inteligencia de negocios (BI)

Consiste en transformar la información cruda del mercado en inteligencia de negocio. Un sistema lo aplica a través de dashboards visuales con el fin de permitir a los usuarios tomar decisiones estratégicas de forma ágil.

Robots.txt

Un archivo de texto, presente en la mayoría de los sitios web, que define las reglas de acceso para crawlers y scrapers.

Inteligencia artificial explicable (XAI)

Conjunto de técnicas (como SHAP) que permiten interpretar las decisiones de modelos de "caja negra". Es crucial para generar confianza y validar los resultados del modelo predictivo.

Objetivo general

Desarrollar una aplicación para comparar precios de la competencia del sector retail de alimentos venezolano, con Web Scraping y Aprendizaje Automático, utilizando inteligencia de negocio para la visualización y análisis de la información obtenida.

Objetivos específicos (I)

La primera fase establece los cimientos tecnológicos y arquitectónicos necesarios para construir una solución robusta y escalable.

(1) Investigación tecnológica

Realizar una revisión bibliográfica de las técnicas de price scraping, sus herramientas asociadas, procesamiento y almacenamiento de datos, aprendizaje automático e inteligencia de negocios, para obtener el marco conceptual de la investigación.

(2) Selección de técnicas

Seleccionar las técnicas de price scraping, algoritmos de aprendizaje automático y herramientas de BI adecuadas para la solución del problema.

(3) Arquitectura modular

Diseñar la arquitectura de software modular de la aplicación para la definición de sus componentes clave y la forma en que interactúan entre sí.

(4) Módulo de scraping

Desarrollar el módulo de price scraping que permita extraer precios de productos específicos de los sitios web de empresas de retail de alimentos seleccionadas en Venezuela.

Objetivo específicos (II)

La segunda fase integra inteligencia avanzada y facilita la visualización estratégica de datos para impulsar decisiones de negocio fundamentadas.

(5) Modelo predictivo

Integrar los algoritmos de aprendizaje automático seleccionados sobre los datos de precios recolectados, para la identificación de patrones significativos en la variación de precios de productos.

(6) Dashboard interactivo

Desarrollar el módulo de visualización y análisis de precios con herramientas de BI, que presente de forma integrada los datos obtenidos mediante price scraping y los resultados del análisis realizado con aprendizaje automático.

(7) Evaluación y validación

Evaluar el rendimiento y precisión de la aplicación para identificar áreas de mejora y sentar una base sólida para desarrollos futuros.

Metodologías aplicadas

La solución integra tres enfoques metodológicos complementarios para asegurar relevancia contextual, desarrollo ágil y manejo científico de datos.

Investigación-acción

Metodología que guió el proyecto a través de ciclos de planificación, acción y reflexión. Esta aproximación iterativa se utilizó para adaptar la solución a los desafíos de un mercado volátil y a la retroalimentación del retailer.



eXtreme Programming (XP)

Desarrollo iterativo y ágil con ciclos cortos, pruebas continuas y adaptación constante a cambios y retroalimentación.

Proceso KDD

Extracción, limpieza y modelado estructurado de datos obtenidos mediante web scraping, siguiendo mejores prácticas científicas.

Arquitectura Funcional del Sistema

La solución integra cuatro módulos interconectados que forman un pipeline completo: desde la captura automatizada de datos hasta la presentación de inteligencia accionable.



Módulo 1: extracción y transformación (ETL)

Web scraping automatizado capture precios en tiempo real de múltiples competidores.



Módulo 2: almacenamiento persistente

PostgreSQL gestiona datos históricos estructurados, proporcionando acceso rápido, escalabilidad y confiabilidad para análisis complejos.



Módulo 3: procesamiento analítico

Modelo de machine learning detecta patrones, realiza predicciones de precios óptimos y genera recomendaciones estratégicas fundamentadas.



Módulo 4: presentación y visualización

Dashboard de BI interactivo presenta insights en tiempo real, facilitando análisis exploratorio y toma de decisiones ágiles.

Arquitectura

Fases del proyecto:

Versión 1 - fundamentos y datos

Construcción del pipeline ETL con web scraping y la capa de persistencia (base de datos) para asegurar un flujo de datos robusto, limpio y confiable.

Versión 3 - interfaz y visualización

Creación del dashboard de inteligencia de negocio y el simulador 'what-if', transformando los resultados del modelo en una herramienta estratégica y accionable para el usuario final.



Versión 2 - inteligencia predictiva

Desarrollo y validación del modelo de machine learning (Random Forest) para analizar los datos extraídos y generar recomendaciones de precios precisas.

Versión 4 - validación integral

Ejecución de pruebas completas de rendimiento, aceptación y validación cualitativa para garantizar la eficiencia, precisión y valor estratégico de la solución final.

Arquitectura

Módulo 1: extracción y transformación (ETL)

Combina dos estrategias complementarias para garantizar una extracción de datos robusta, eficiente y adaptable a cualquier tipo de sitio web de e-commerce.

Estrategia 1: automatización de navegador

Uso de Selenium para sitios web dinámicos y complejos que dependen de JavaScript.

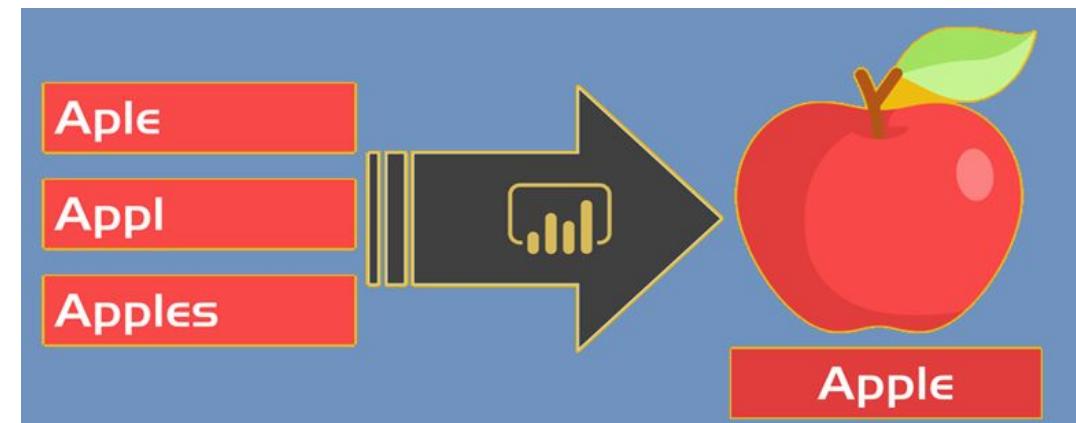
Estrategia 2: ingeniería inversa de APIs

Consumo directo de APIs internas de los e-commerce.



Preprocesamiento y normalización

De igual forma, la técnica principal para preparar los datos fue la homologación de productos mediante 'Fuzzy String Matching'. Este algoritmo fue necesaria para estandarizar la información, permitiendo corregir errores tipográficos, unificar formatos y manejar inconsistencias.

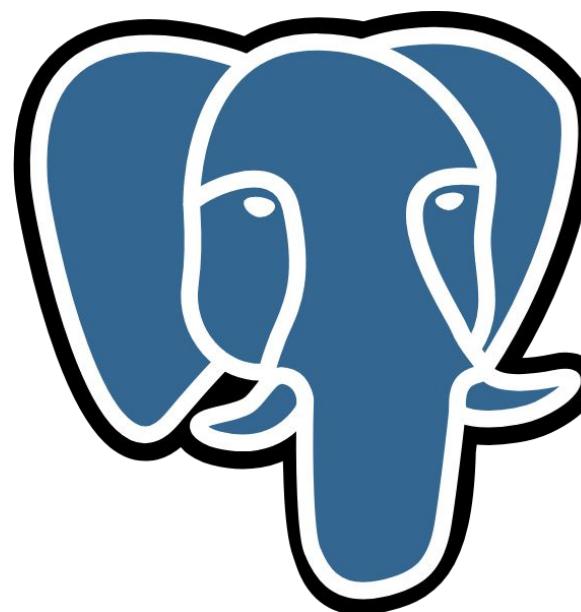


Arquitectura

Módulo 2: almacenamiento persistente

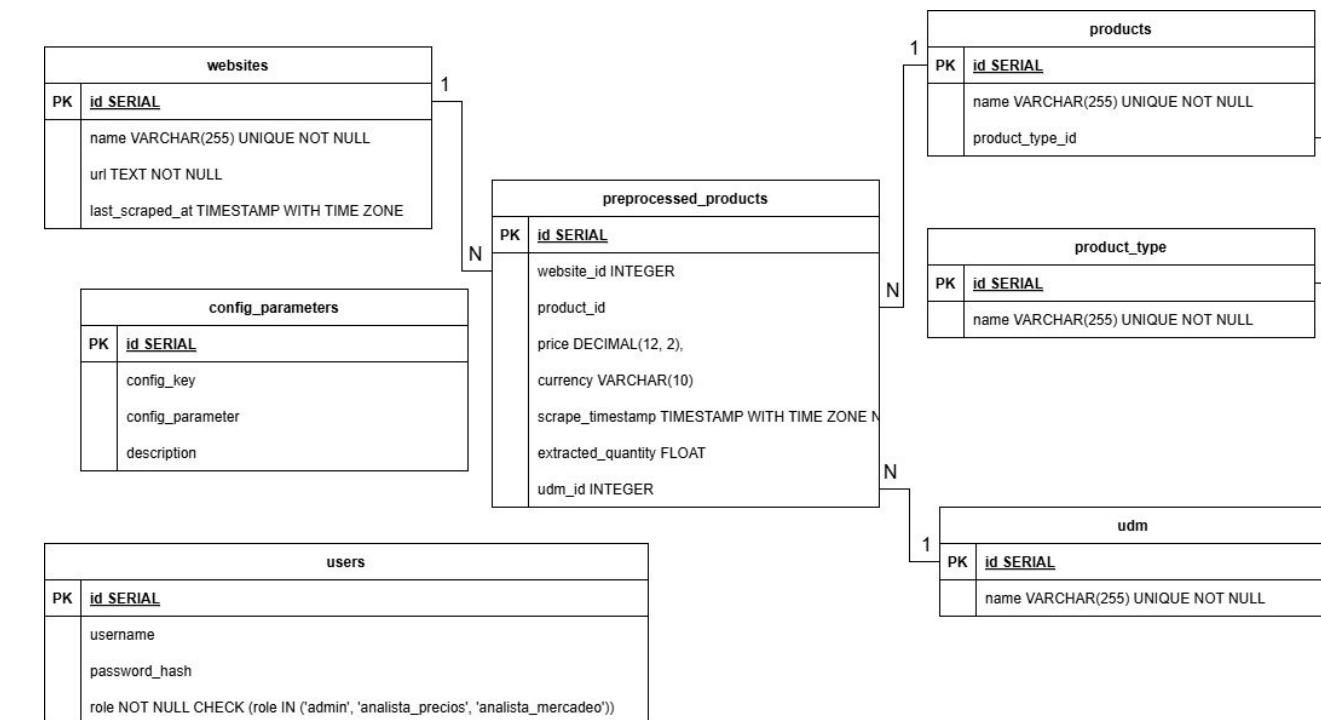
Tecnología utilizada

Se seleccionó PostgreSQL 14 como sistema de gestión de bases de datos relacionales (RDBMS) por su robustez, y su excelente integración con Python, facilitando un desarrollo eficiente y escalable.



Entidades principales

- websites: Representa a los retailers monitoreados.
- products: Catálogo maestro de productos homologados.
- product_type, udm: Tablas catálogo para estandarizar tipos de producto y unidades de medida.
- preprocessed_products: Tabla transaccional que almacena cada dato de precio recolectado, constituyendo el histórico central.
- config_parameters: Almacena parámetros de configuración del sistema para evitar el uso de parámetros estáticos en el código.
- users: Proporciona el grupo de usuarios disponibles en la aplicación.



Arquitectura

Módulo 3: procesamiento analítico

Nuestro modelo predictivo transforma datos históricos en recomendaciones de precios óptimos mediante un flujo de trabajo científico y validado.



Dataset

Datos históricos propios y de la competencia obtenidos mediante web scraping.

Algoritmo seleccionado

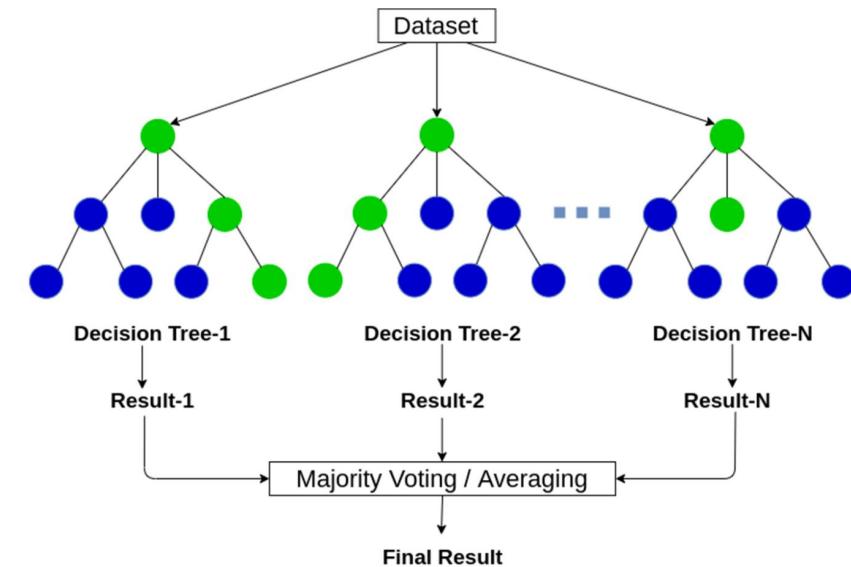
Random Forest Regressor - Precisión y estabilidad en mercados volátiles.

Predictión

Recomendación de precio óptimo adaptada al mercado.

- ❑ **Ventaja clave:** Capaz de modelar relaciones no lineales y complejas del mercado, identificando patrones que los métodos tradicionales no pueden detectar.

Random Forest



Arquitectura

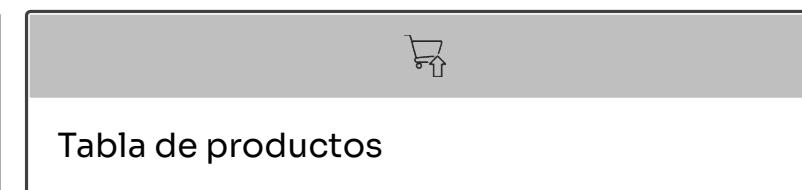
Módulo 4: presentación y visualización

Tecnología principal

Plotly: Utilizada como librería principal para la creación de aplicaciones web con gráficos interactivos de alta calidad.

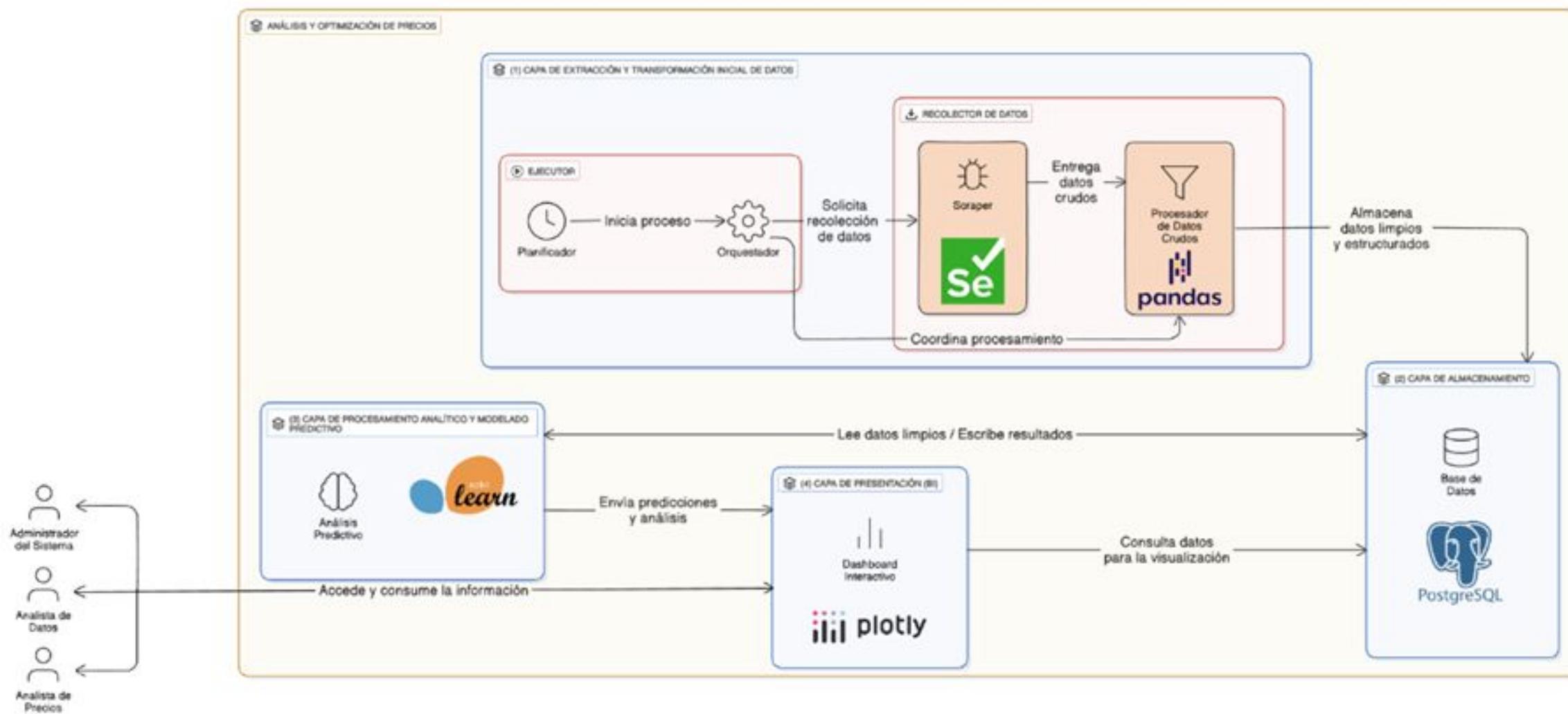
Tipos de gráficos: Soporte para la visualización de datos mediante gráficos de líneas, barras y diagramas de violín, adaptados a las necesidades del análisis de precios.

Funcionalidades implementadas



Arquitectura

Diagrama de arquitectura



Resultados (Implementación)

Historias de usuario

Analista de precios

-  **HU-01:** Ver una tabla que compare mi precio, el de la competencia y el precio recomendado.
-  **HU-02:** Visualizar un ranking de los productos con mayor oportunidad de ajuste de precio.
-  **HU-03:** Simular el impacto en el precio recomendado si cambio los precios de mis competidores.
-  **HU-05:** Comparar la distribución general de precios de una categoría contra la competencia.

Analista de datos

-  **HU-04:** Ver la evolución histórica del precio de un producto, comparando mi empresa con la competencia.
-  **HU-06:** Filtrar todos los análisis por múltiples criterios (producto, fecha, etc).
-  **HU-07:** Acceder a una tabla con todos los datos históricos recolectados.
-  **HU-08:** Exportar los datos filtrados a un archivo CSV.

Administrador

-  **HU-09:** Que el sistema recolecte los datos de la competencia de forma automática y periódica.

Resultados (Implementación)

Dashboard de análisis de precios:

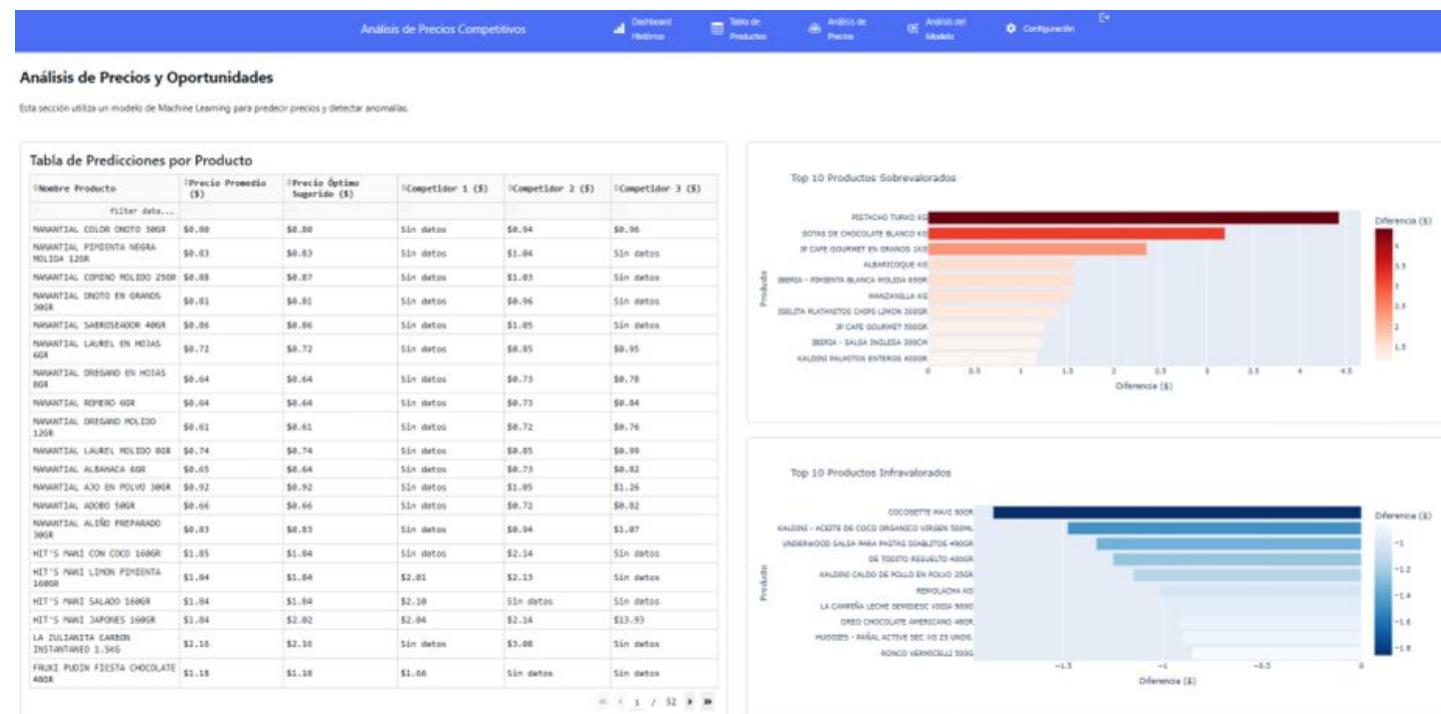
Centraliza la información para la toma de decisiones estratégicas, permitiendo a los equipos actuar con rapidez y confianza.

Tabla comparativa

Visualiza precio actual vs. competencia vs. precio recomendado en una matriz clara y accionable que facilita la identificación inmediata de oportunidades.

Identificación de oportunidades

Gráficos de barras que destacan las mayores oportunidades de optimización, tanto al alza como a la baja, priorizando dónde enfocar esfuerzos.



Resultados (Implementación)

Planificación proactiva: simulador "what-if"

Permite a los analistas anticipar movimientos del mercado y probar el impacto de diferentes escenarios de precios antes de que ocurran en la realidad.

The screenshot shows a product analysis dashboard with a table of products and a price prediction overlay. The table includes columns for product name, current price, and suggested optimal price. Below the table is a 'Simulator de Precio Óptimo' section where users can input current price and competitor prices to predict optimal prices.

Nombre	Precio Actual (\$)	Precio Óptimo Sugerido (\$)
SAVOY POSTRES OSCURO 55% 200GR	\$4.66	\$4.45
SAVOY POSTRES OSCURO 40% 200GR	\$4.67	\$4.67
NESTLE - CHOCOLATE 27GR	\$1.17	\$1.20
TORONTO CHOCOLATE BOLSA 125GR	\$5.67	\$5.62
PAN MEZCLA PARA CACHAPA 500 GR	\$1.95	\$1.94
SAMBÀ FRESA 32GR	\$0.79	\$0.87



Modificación de variables

El analista puede ajustar variables del mercado en tiempo real, como cambios en precios de competidores.



Impacto instantáneo

Visualiza inmediatamente cómo cada cambio afecta la recomendación del modelo, permitiendo evaluar múltiples escenarios antes de tomar decisiones.



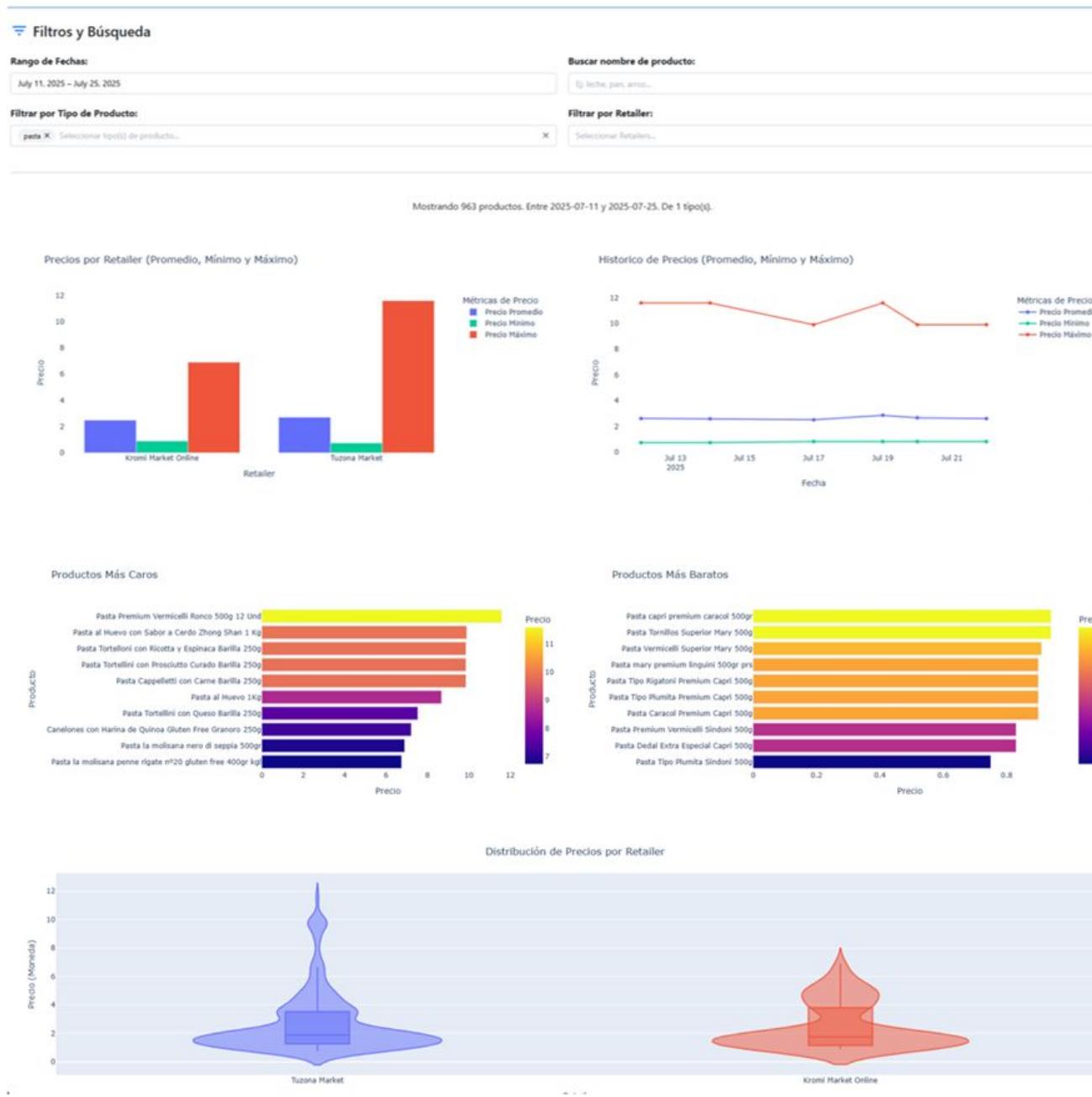
Planificación estratégica

Transforma datos en estrategia, permitiendo preparar respuestas a diferentes condiciones de mercado antes de que ocurran.

Resultados (Implementación)

Dashboard histórico

El análisis histórico competitivo proporciona contexto y perspectiva sobre las dinámicas del mercado a lo largo del tiempo.



Evolución de precios

Visualiza tendencias y reacciones del mercado a lo largo del tiempo, identificando patrones estacionales y estrategias de competidores.



Distribución y estrategia

Gráficos de violín comparan el posicionamiento de precios entre retailers, revelando quién es consistentemente más caro o más barato.

Resultados (Implementación)

Dashboard de productos

Ofrece una tabla completa con el historial de precios de todos los productos, comparados entre la competencia

Funcionalidades clave

Vista tabular completa

Accede a todos los datos de la tabla preprocessed_products.

Filtros dinámicos

Organiza la información por fecha, producto y retailer.

Integración con excel

Facilita la exportación de datos para análisis avanzados y personalizados.

The screenshot shows a user interface for a product dashboard. At the top, there are several filter options: 'Filtros y Búsqueda' (Filters and Search) with date range 'July 11, 2025 – July 25, 2025', a search bar 'Buscar nombre de producto:' with placeholder 'Ej. leche, pan, arroz...', and dropdowns for 'Filtrar por Tipo de Producto:' (Select product type) and 'Filtrar por Retailer:' (Select retailer). Below the filters is a section titled 'Lista de Productos Filtrados' (List of Filtered Products) with a sub-section 'Mostrando 30323 productos.' (Showing 30323 products). A green button 'Exportar a Excel' (Export to Excel) is visible. The main area displays a table with columns: Nombre (Name), Precio (Price), Moneda (Currency), Fecha (Date), Cantidad (Quantity), Fuente (Source), and CUDI. The table lists various products like 'Detergente en Polvo Original Purex 400g', 'Infusión Paraíso Azul sin Azúcar LucyA 630g', etc., with their respective details.

Nombre	Precio	Moneda	Fecha	Cantidad	Fuente	CUDI
filter data...						
Detergente en Polvo Original Purex 400g	2.1	BSD	22/07/2025	400	Tuzona Market	gramos
Infusión Paraíso Azul sin Azúcar LucyA 630g	16	BSD	22/07/2025	630	Tuzona Market	gramos
Salsa Condimentada Iberia 300ml	5.52	BSD	22/07/2025	300	Tuzona Market	millilitres
Salsa de Chile Estilo Guulin Lee Kum Kee 230g	12.2	BSD	22/07/2025	230	Tuzona Market	gramos
Salsa Oscura de Anguila para Parrilla 200ml	6.35	BSD	22/07/2025	200	Tuzona Market	millilitres
Vinagre para Sushi Kong Yen 100ml	11.63	BSD	22/07/2025	300	Tuzona Market	millilitres

Resultados (Implementación)

Dashboard de análisis del modelo

Análisis exploratorio de datos (EDA)

- Histogramas para distribución y frecuencia de variables clave.
- Matriz de correlación entre variables del modelo.
- Tabla descriptiva con métricas estadísticas (media, desviación estándar, cuartiles).

Evaluación del rendimiento

- Gráfico de dispersión (predichos vs reales).
- Histograma de residuos para identificar sesgos.
- Gráfico Q-Q para verificar distribución normal de errores.

Interpretabilidad del modelo (explainable AI)

- Gráfico de resumen SHAP para impacto global de características.
- Gráfico de dependencia parcial para análisis de variables individuales.
- Desglose de contribución por variable para predicciones específicas.

Análisis Profundo del Modelo de Machine Learning

Esta sección ofrece una vista detallada del conjunto de datos, el rendimiento del modelo y su interpretabilidad.

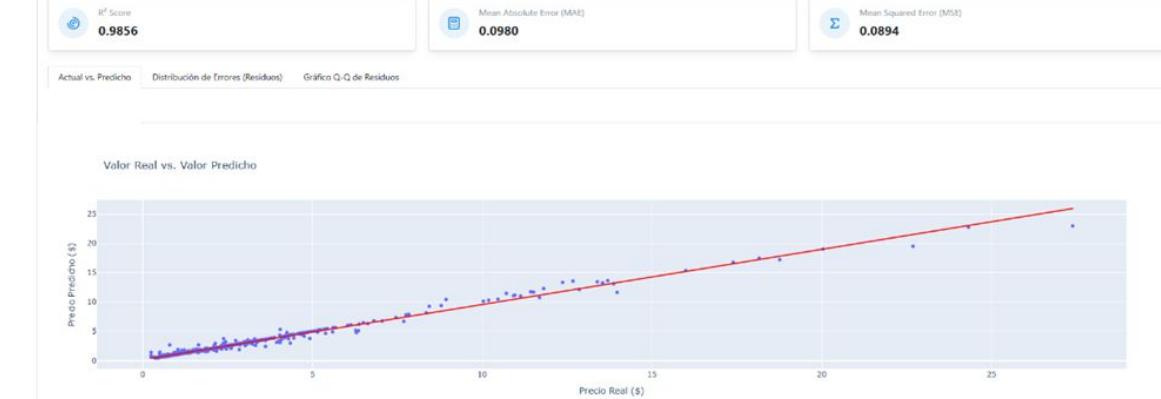
Analís Exploratorio de Datos (EDA)

Descripción del Dataset | Distribuciones de Variables | Matriz de Correlación

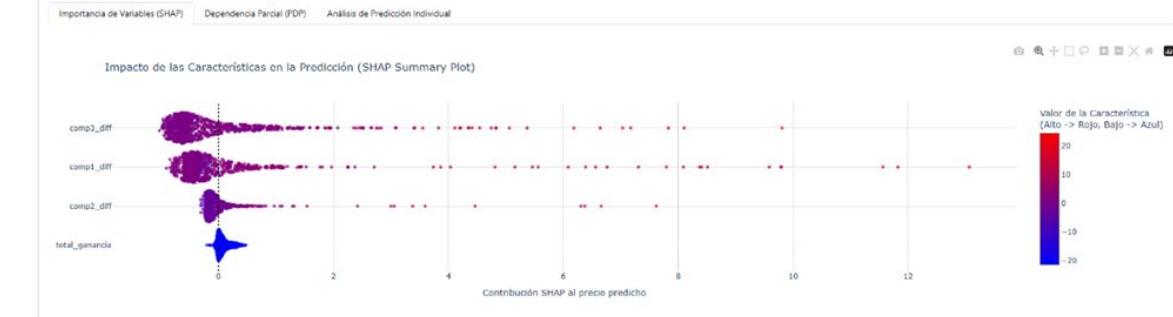
Descargar Dataset

	total_ganancia	comp1_diff	comp2_diff	comp3_diff	precio_mes_actual
index	1023	1023	1023	1023	1023
count	1402,4199	0,8337	0,5098	1,2695	2,3656
mean	5285,8432	2,9157	2,3187	2,6242	2,4965
std	4	-26,5173	-18,2867	-21,2819	0,245
min	190,05	-0,39	-0,3004	-0,1647	1,1147
25%	439,24	0,6532	-0,048	1,0627	1,7075
50%	1094,035	1,8462	1,0562	2,1	2,6858
75%	121617,6	27,3895	27,3895	24,3204	27,3895
max					

Evaluación del Rendimiento del Modelo



Interpretabilidad del Modelo (Explainable AI)



Resultados (Implementación)

Panel de administración

Panel de Administración

Gestión de Usuarios
Crear nuevos usuarios y asignarles un rol en el sistema.

Nombre de Usuario *

Contraseña *

Role del Usuario *

Crear Usuario

Editar o Eliminar Usuario
Seleccionar Usuario

Nuevo Nombre de Usuario (Opcional)

Nueva Contraseña (Opcional)

Actualizar Usuario **Eliminar Usuario**

Parámetros del Sistema
Visualiza, edita y crea parámetros de configuración del sistema.

Parámetros Actuales (haz clic en una fila para editar):

Clave	Valor	Descripción	Última Actualización
INITIAL_ADMIN_CREATED	true	Bandera para asegurar que el usuario admin inicial se cree solo una vez.	2025-10-14 21:28:24
initial_product_load_completed	true	Flag booleano ('true'/'false') que indica si la carga de datos iniciales en la tabla 'products' ya se ha completado.	2025-08-04 22:34:38
MAIN_JOB_SCHEDULE_TIME	18:50	Hora de ejecución (HH:MM) para el job principal del scheduler.	2025-08-04 22:34:29

Clave del Parámetro *

Valor del Parámetro *

Descripción (Opcional)

Limpiar Campos **Guardar Cambios**

Funcionalidades principales

- Gestión de usuarios del sistema.
- Administración de parámetros del sistema.
- Control de configuraciones operativas.
- Supervisión de accesos y permisos.

Características clave

- Interfaz administrativa centralizada.
- Control total sobre la configuración del sistema.
- Gestión de credenciales y roles de usuario.
- Monitoreo de parámetros operativos.

Resultados (Validación)

Interpretabilidad del modelo: análisis PDP y SHAP

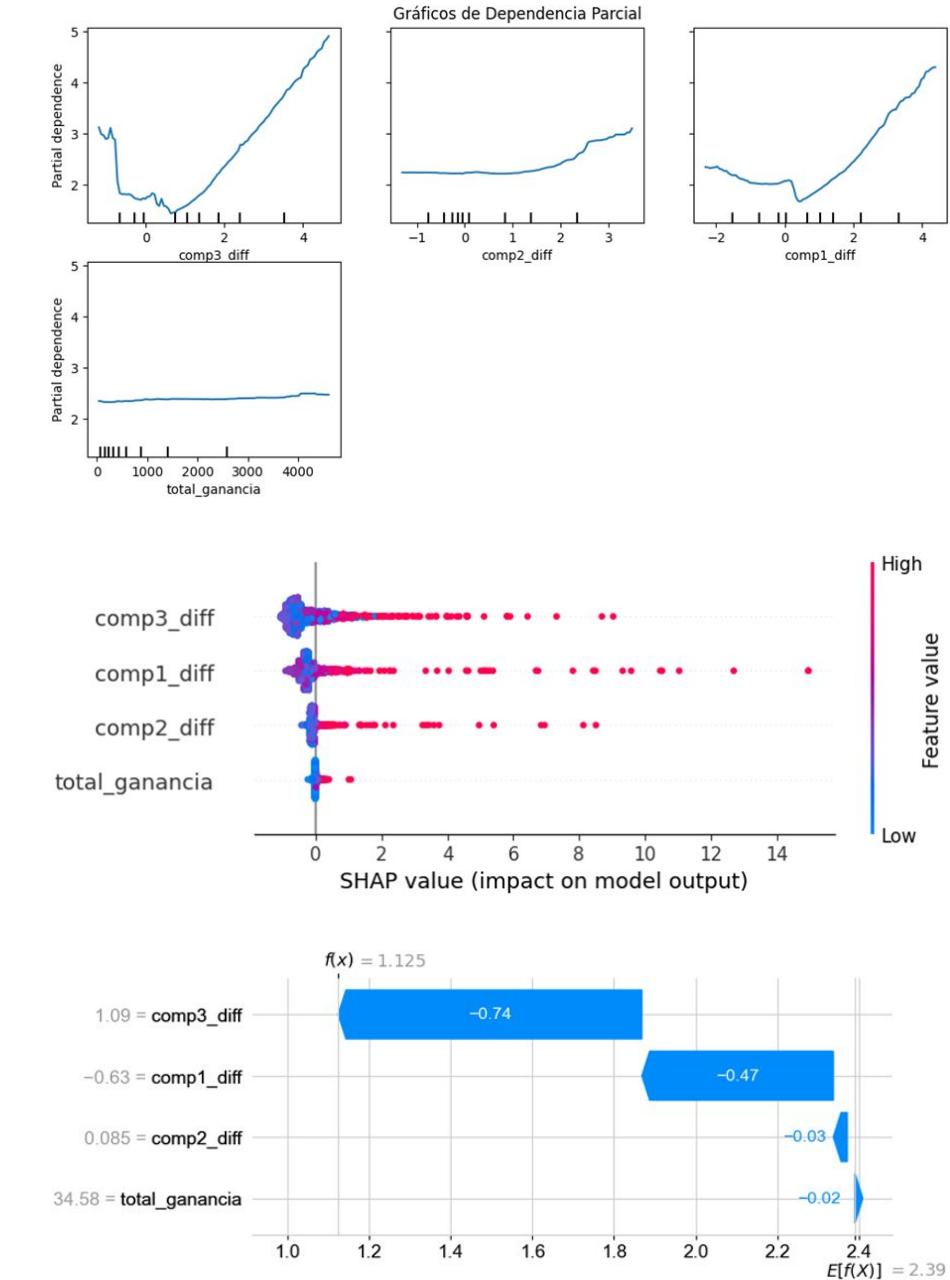
Hallazgos del análisis de dependencia parcial (PDP)

- Relación convexa y no lineal entre diferencias de precios (`comp3_diff` y `comp1_diff`) y la predicción.
- Cuando la empresa ofrece precios más baratos, el modelo sugiere un precio más alto, revelando una oportunidad de optimización.
- Curvas planas para `comp2_diff` y `total_ganancia`, indicando una capacidad predictiva limitada de estas variables.

Cuantificación con SHAP

`comp3_diff` y `comp1_diff` son las variables más influyentes en el modelo.

- Existe una correlación positiva directa: valores altos en estas diferencias impulsan la predicción al alza.
- Permite un análisis granular, explicando la contribución de cada característica a nivel de predicción individual.



Resultados (Validación)

Precisión y confiabilidad del modelo

Las métricas de rendimiento validan la capacidad predictiva del modelo, asegurando que las decisiones se basen en recomendaciones fiables y basadas en evidencia.

0.98

R² (Coeficiente de determinación)

El modelo explica el **98% de la variabilidad** de los precios del mercado. Un ajuste excepcionalmente alto que demuestra la robustez del algoritmo.

0.15

MAE (Error absoluto medio)

En promedio, las predicciones se desvían sólo **0.15 unidades monetarias** del precio real. Alta precisión para decisiones de negocio con confianza.

Conclusiones clave

Se validó una solución integral de ciencia de datos para el retail venezolano, con precisión predictiva excepcional y capacidad transformadora para la toma de decisiones estratégica.



Viabilidad demostrada

Se validó la viabilidad técnica de una solución end-to-end, desde captura de datos con Web Scraping hasta recomendaciones inteligentes para optimización de precios.



Precisión predictiva del 98%

El modelo de Machine Learning (Random Forest) alcanzó una precisión R^2 de 98%, posicionándose como herramienta confiable para decisiones críticas de pricing.

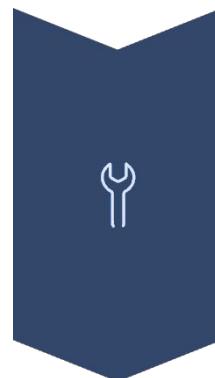


Valor estratégico real

El dashboard interactivo transforma datos en inteligencia accionable, permitiendo toma de decisiones proactiva y basada en evidencia para ventaja competitiva.

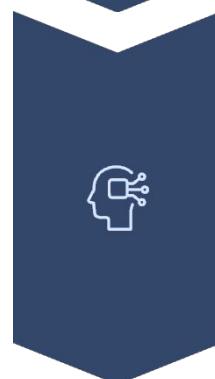
Próximos pasos

Estrategia de evolución para maximizar el impacto: mejorar la robustez de extracción, enriquecer modelos predictivos e integrar completamente con ecosistemas retailers.



Fortalecer extracción de datos

Desarrollar módulo de auto-recuperación para web scraping que se adapte dinámicamente a cambios en estructuras web, garantizando continuidad operativa y confiabilidad del flujo de datos.



Enriquecer el modelo predictivo

Incorporar variables adicionales: promociones, niveles de stock, indicadores macroeconómicos y explorar modelos de series temporales para predicciones más sofisticadas y contextualizadas.



Integración total de ecosistema

Conectar solución con sistemas internos del retailer (ERP, inventario, punto de venta) creando plataforma 360° de gestión de precios inteligente y automatizada.

Muchas gracias

¡Gracias por su atención! ¿Preguntas?

Referencias bibliográficas

- Abodayeh, A., Shihadeh, L., Hejazi, R., Latif, R., & Najjar, W. (2023). Web Scraping for Data Analytics: A BeautifulSoup Implementation. 2023 IEEE/WDS/FRJK International Conference. <https://doi.org/10.1109/WDS-FRJK1971.2023.00235>
- Agile Proceddings (2023) Agile Processes in Software Engineering and Extreme Programming: 24th International Conference on Agile Software Development, XP 2023, Amsterdam, The Netherlands, June 13–16, 2023, Proceedings. (2023). Suiza: Springer Nature Switzerland.
- Almohammadi, B. O. (2019). How Business Intelligence Can Help You to Better Understand Your Customers. ResearchGate. Recuperado de https://www.researchgate.net/publication/338326425_How_Business_Intelligence_Can_Help_You_to_Better_Understand_Your_Customers
- Boring Owl. (2023, Mayo 31). The Power of Web Scraping for Business Intelligence. Software House Boring Owl. Recuperado de <https://boringowl.io/en/blog/why-web-scraping-is-important-for-business-intelligence>
- Bradbury, H. (2015). The SAGE handbook of action research (H. Bradbury-Huang, Ed.; 3a ed.). SAGE Publications.
- Chang, C.-Y., & He, X. (2025). The Liabilities of Robots.txt. arXiv preprint arXiv:2503.06035. <https://arxiv.org/abs/2503.06035>
- Cobo, L., Benítez Baldión, S., Perdomo González, J. S., & Novoa Mendoza, H. A. (2024). Web scraping en supermercados para el seguimiento de precios de la cesta básica alimentaria. Ciencia e ingeniería, 12(1), e14533404. <https://doi.org/10.5281/zenodo.14533404>
- D'Souza, M., Agrawal, D., Desai, S., & Joshi, F. (2024). Web Scraping based Product Comparison Model for E-Commerce Website. Journal of Emerging Technologies and Innovative Research (JETIR).
- Ernest T. Stringer (2007) Action research third edition (3rd. ed.). Sage publications
- Han, J., Pei, J., & Tong, H. (2022). Data mining: Concepts and techniques (4a ed.). Morgan Kaufmann.

Referencias bibliográficas

- Intelligence Node. (s.f.). The Importance of Competitor Price Comparison. Recuperado de <https://www.intelligencenode.com/blog/the-importance-of-competitor-price-comparison/>
- Jean McNiff, Jack Whitehead. (2002). Action Research: Principles and Practice (2da ed.)
- Joyanes Aguilar, L. (2020). Inteligencia de negocios y analítica de datos. España: Marcombo.
- Kaizen Institute. (s.f.). Optimizing Business Intelligence for Strategic Advantage. Recuperado de <https://kaizen.com/insights/business-intelligence-strategic-advantage/>
- Kent Beck, Cynthia Andres. (2004). Extreme Programming Explained: Embrace Change, Second Edition
- Khder, M. A. (2021). Web Scraping or Web Crawling; State of Art, Techniques, Approaches and Application. International Journal of Advanced Soft Computing and Applications.
- Nagle, T. T., & Muller, G. (2018). The strategy and tactics of pricing: A guide to growing more profitably (6a ed.). Routledge.
- Nouel, F. & Rodríguez, R. (2024). Adaptación al cambio tecnológico: Diagnóstico de la transformación digital empresarial en Venezuela. Disponible en: <https://www.researchgate.net/publication/385659647 ADAPTACION AL CAMBIO TECNOLOGICO DIAGNOSTICO DE LA TRANSFORMACION DIGITAL EMPRESARIAL EN VENEZUELA>
- Nuñez Cartolin, C. A. (2021). Business Intelligence y su impacto en la productividad del proceso de toma de decisiones de la alta gerencia en la empresa Newocean Technology S.A.C. (Trabajo de suficiencia profesional). Universidad Privada del Norte, Lima, Perú.

Referencias bibliográficas

- Papadopoulou, M. I. (2023). Understanding Business Intelligence and Dashboard Creation: An In-depth Analysis and Conceptual Framework (GRI-2023-41843) [Graduate thesis, Aristotle University of Thessaloniki, School of Sciences, Department of Informatics].
- Pertuz, C. M. P. (2022). Aprendizaje automático y profundo en python. Ra-Ma Editorial. Observatorio Venezolano de Finanzas (OVF). (Marzo 2025). <https://observatoriodefinanzas.com/inflacion-en-venezuela-febrero-2025-con-un-12-8-mensual-y-117-interanual/>
- Price2Spy. (2024, nov). What is Price Scraping? Price2Spy Blog. Recuperado de <https://www.price2spy.com/blog/price-scraping/>
- Torres Benitez, G. E. (2020). Modelo de inteligencia de negocios como herramienta para la toma de decisiones en el ámbito gerencial: Caso Iterauto Primum C.A., Naguanagua Edo. Carabobo 2020 [Trabajo de Grado de Maestría]. Universidad Jose Antonio Paez.
- Valecillos, O. (2019). Desarrollo de un sistema de recomendaciones para un sitio de Comercio Electrónico (Trabajo Especial de Grado). Universidad Central de Venezuela, Caracas.
- Venigandla, K., Vemuri, N., Thaneeru, N., & Tatikonda, V. M. (2023). Leveraging AI-Enhanced Robotic Process Automation for Retail Pricing Optimization: A Comprehensive Analysis. Journal Of Knowledge Learning and Science Technology ISSN 2959-6386 (Online), 2(2), 361-370. <https://doi.org/10.60087/jklst.vol2.n2.p370>
- Singh, H. (2023). Retail price optimization. Kaggle. <https://www.kaggle.com/code/harhsingh2209/retail-price-optimization/notebook>
- López Martín, P. (2023). Técnicas de machine learning e interpretabilidad aplicadas al mercado inmobiliario [Trabajo Fin de Grado, Universidad Complutense de Madrid].

Referencias bibliográficas

- Dib, L., & Capus, L. (2025). Classifying XAI Methods to Resolve Conceptual Ambiguity. *Technologies*, 13(9), 390. <https://doi.org/10.3390/technologies13090390>
- Brown, M. A., Gruen, A., Maldoff, G., Messing, S., Sanderson, Z., & Zimmer, M. (2024, oct). Web scraping for research: Legal, ethical, institutional, and scientific considerations. <https://arxiv.org/abs/2410.23432>
- Lakshmi, A. (2025, febrero). Ethical web scraping: A practical guide to responsible data collection. Scraperapi. <https://www.scrapingapi.com/web-scraping/ethical/>
- R cuadrado en R: interpretación y cálculo. (2024, julio). IONOS Digital Guide. <https://www.ionos.com/es-us/digitalguide/paginas-web/desarrollo-web/r-squared-in-r/>
- Lhotanová, K. (2025, febrero). How to reverse engineer website APIs. Apify Blog. <https://blog.apify.com/reverse-engineer-apis/>