

**Evidencia de aprendizaje 1. Automatización de la Recolección de Datos Históricos  
de un Indicador Económico**

Adriana María Aguilar Vilorio

Edwin Bernardo Villa Sánchez

**Código de grupo:**

PREICA2501B020128

**Docente:**

Andrés Felipe Callejas

**Asignatura:**

Proyecto integrado V

Programa Ingeniería de Software y Datos

Facultad Ingenierías y Ciencias Agropecuarias

Institución Universitaria Digital de Antioquia

2025

## Tabla de Contenido

Resumen	3
Introducción	5
Objetivo General	6
Objetivos Específicos	7
KPIs para el análisis del indicador 6A=F	8
Metodología	9
Resultados	12
Bibliografía	13

## Resumen

Este informe documenta el desarrollo e implementación de una solución automatizada para la recolección, procesamiento y almacenamiento continuo de datos históricos del indicador financiero 6A=F, correspondiente al futuro del dólar australiano (AUD/USD). Este activo, listado en los mercados financieros internacionales y consultado desde Yahoo Finanzas, representa un instrumento fundamental para el análisis macroeconómico, la toma de decisiones en trading y la gestión de riesgos cambiarios.

El proyecto fue concebido para eliminar la intervención manual en la recolección de información bursátil, garantizando persistencia, trazabilidad, integridad y automatización del proceso mediante tecnologías modernas como Python, GitHub Actions y GitHub Code Spaces. El núcleo de la solución se basa en un sistema de scraping construido con requests y BeautifulSoup, que permite extraer datos relevantes como: fecha, apertura, cierre, máximos, mínimos, volumen y cierre ajustado.

Una vez extraída la información, esta es limpiada y transformada en un DataFrame estructurado con pandas, aplicando conversiones de fechas, normalización de valores numéricos y segmentación temporal (año, mes, día). El flujo de datos está encapsulado bajo un enfoque de programación orientada a objetos (POO), donde clases como Collector y Logger permiten modularizar responsabilidades, mejorar el mantenimiento del código y facilitar futuras extensiones a otros indicadores económicos.

Además, se diseñó un flujo de integración y despliegue continuo (CI/CD) mediante GitHub Actions, el cual se ejecuta automáticamente tras cada actualización del repositorio. Este flujo prepara el entorno, ejecuta el script de extracción y almacenamiento, y realiza un commit con los nuevos datos al historial versionado. Esto garantiza que el archivo `historical.csv` contenga información actualizada, libre de duplicados y lista para análisis financiero.

El entorno completo fue diseñado para ejecutarse sin fricciones locales, utilizando GitHub Codespaces como entorno de desarrollo remoto. Este enfoque facilita el trabajo colaborativo, la portabilidad del proyecto y la reducción de errores por diferencias de configuración entre equipos de desarrollo.

En resumen, se logró construir una solución para el seguimiento automatizado del comportamiento del dólar australiano en el tiempo, alineada con buenas prácticas de ingeniería de software y análisis de datos. Este sistema no solo contribuye a la eficiencia en la recolección de datos financieros, sino que también abre el camino a análisis predictivos y visualizaciones estadísticas orientadas a la toma de decisiones en el ámbito económico y de inversión.

## Introducción

El presente informe se centra en el análisis automatizado del indicador económico **6A=F**, correspondiente al contrato de futuros del **dólar australiano (AUD)** listado en la plataforma de **Yahoo Finanzas**. Este indicador refleja la evolución del tipo de cambio entre el AUD y el USD a través de derivados financieros negociados en los mercados internacionales. La cotización 6A=F proporciona información diaria clave como precios de apertura, cierre, máximos, mínimos, volumen de transacciones y cierre ajustado.

El futuro del dólar australiano ha sido parte del mercado bursátil por varias décadas y es un instrumento crucial para la cobertura de riesgos (hedging), inversión especulativa y análisis macroeconómico. Este tipo de activo es utilizado por analistas financieros, entidades comerciales y gobiernos para evaluar la fortaleza relativa de la economía australiana frente al dólar estadounidense.

La solución técnica desarrollada en este proyecto permite recolectar datos históricos de este indicador de manera automatizada, segura y reproducible. Además, integra prácticas avanzadas de programación, versión de código y despliegue continuo, lo que facilita su escalabilidad y adaptación a otros activos financieros similares.

## **Objetivo General**

Automatizar la recolección continua de datos históricos de un indicador económico, garantizando su persistencia y trazabilidad en un entorno de control de versiones y documentación inicial clara.

### Objetivos Específicos

- Obtener datos históricos confiables del indicador 6A=F desde la fuente oficial Yahoo Finanzas, utilizando técnicas de scraping con Python.
- Preprocesar los datos recolectados, transformando las series temporales a un formato estructurado y normalizado, eliminando duplicados y valores atípicos.
- Diseñar un sistema orientado a objetos reutilizables, que encapsule los procesos de recolección, transformación, validación y persistencia de datos, facilitando su extensión a otros indicadores.
- Implementar un sistema de logging configurable, que permita auditar cada ejecución y detectar posibles errores en el flujo de trabajo.
- Automatizar la ejecución del pipeline de datos con GitHub Actions, garantizando la actualización continua, trazabilidad e integración directa al repositorio.
- Almacenar los datos históricos en un archivo consolidado (.csv) que mantenga la integridad y permita su uso en análisis posteriores.
- Visualizar indicadores clave del comportamiento histórico del AUD/USD, como tendencias mensuales, volatilidad promedio y volumen negociado, a través de reportes estadísticos y representaciones gráficas.

### KPIs para el análisis del indicador 6A=F

- Evolución del Precio de Cierre Diario
  - Para qué sirve: Permite observar el comportamiento del precio en el muy corto plazo, identificando la tendencia y volatilidad más recientes.
  - Cómo se calcula (Métrica): Precio de Cierre ('Close') de cada día.
  - Presentación en el Dashboard: Gráfico de líneas "Precio de Cierre Diario - 2025".
  
- Tendencia de Mediano Plazo (Media Móvil)
  - Para qué sirve: Suaviza la volatilidad del precio diario para identificar la dirección fundamental del mercado (tendencia alcista, bajista o lateral) y los principales ciclos económicos.
  - Cómo se calcula (Métrica): Media Móvil de 30 días = Suma de los precios de cierre de los últimos 30 días / 30.
  - Presentación en el Dashboard: Gráfico de líneas "Media Móvil de 30 Días", comparando el precio con su media móvil.
  
- Volatilidad Histórica Mensual
  - Para qué sirve: Visualiza el grado de riesgo e incertidumbre del mercado en un periodo específico (mes). Picos altos indican pánico o euforia; valles bajos indican calma.
  - Cómo se calcula (Métrica): Calculada como la desviación estándar de los precios de cierre diarios dentro de cada mes.



- Presentación en el Dashboard: Gráfico de líneas "Volatilidad Mensual".
  
- Desviación Estándar Mensual
  - Para qué sirve: Mide cuantitativamente la dispersión estadística de los precios de cierre diarios con respecto a su precio promedio durante cada mes. Es la medida técnica que define la volatilidad.
  - Cómo se calcula (Métrica):  $\sigma = \sqrt{[\Sigma(x_i - \mu)^2 / N]}$ , donde  $x_i$  es cada precio de cierre del mes,  $\mu$  es el precio promedio del mes y  $N$  es el número de días.
  - Presentación en el Dashboard: Gráfico de líneas "Desviación Estándar Mensual". (Nota: Este gráfico visualiza la misma métrica que "Volatilidad Mensual", presentado como un elemento separado).
  
- Rendimiento Acumulado de la Inversión (Buy & Hold)
  - Para qué sirve: Muestra cuál habría sido el rendimiento total de una inversión inicial si se hubiera comprado el activo al principio del período y se hubiera mantenido hasta la fecha. Mide la rentabilidad a largo plazo.
  - Cómo se calcula (Métrica): Producto acumulado de  $(1 + \text{el retorno porcentual diario})$ .
  - Presentación en el Dashboard: Gráfico de líneas "Retorno Acumulado Desde el Inicio".

- Identificación de Eventos de Crisis (Variaciones Extremas)
  - Para qué sirve: Identifica con precisión las fechas exactas de pánico o euforia extrema, permitiendo conectar los picos de volatilidad con eventos macroeconómicos concretos.
  - Cómo se calcula (Métrica): Conteo de días donde el cambio porcentual diario fue mayor a +5% o menor a -5%.
  - Presentación en el Dashboard: Tabla "Días con Variaciones Extremas ( $\pm 5\%$ )".

## Metodología

### Selección del Indicador

Se eligió el indicador 6A=F, que representa el precio del futuro del dólar australiano, como fuente principal de datos. La URL base utilizada fue:

<https://es.finance.yahoo.com/quote/6A=F/>

### Recolección de Datos

Se emplearon las bibliotecas requests y BeautifulSoup para extraer la tabla de datos históricos directamente del HTML de la página web. El contenido extraído incluye: fecha, apertura, máximo, mínimo, cierre, cierre ajustado y volumen.

### Limpieza y Estructuración

Los datos se transforman en un DataFrame de pandas, normalizando los valores numéricos y realizando una conversión precisa de fechas al formato datetime. Además, se extraen las componentes de año, mes y día para facilitar futuros análisis.

### Diseño Orientado a Objetos

La clase Collector encapsula el proceso de descarga, validación y guardado de datos. Esta clase es fácilmente reutilizable y extensible para futuros indicadores. Por otro lado, la clase Logger gestiona los registros de ejecución, errores y eventos del sistema, con distintos niveles de severidad (INFO, ERROR, DEBUG), y formato estandarizado. El archivo main.py sirve como punto de entrada y orquestación del flujo de datos.

## Estructura del Repositorio

El repositorio sigue una estructura modular clara:

```
<PROYECTOINTEGRADOV_2025>/
├── .github/
│   └── workflows/
│       └── update_data.yml                # Automatización con GitHub Actions
├── src/
│   └── <indicador_economico>/
│       ├── static/
│       │   ├── data/
│       │   │   ├── dolar_data_enricher.csv    # Dataset enriquecido
│       │   │   ├── historical.csv             # Dataset generado
│       │   │   └── predicciones.csv           # Resultados de predicciones generadas
│       │   ├── models/
│       │   │   └── model.pkl                  # Modelo entrenado serializado
│       │   └── reports/
│       │       ├── Dashboard.pdf              # Reporte PDF con capturas de gráficos
│       │       └── metricas.txt               # Justificación de métricas del modelo
│       ├── collector.py                      # Clase principal para recolección de datos
│       ├── dashboard.py                     # Dashboard con visualización de datos históricos
│       ├── dashboard_prediccion.py          # Dashboard con visualización de predicciones
│       ├── enricher.py                      # Script para enriquecimiento y procesamiento de datos
│       ├── logger.py                        # Configuración y manejo de logs
│       ├── main.py                          # Punto de entrada del programa
│       └── modeller.py                      # Entrenamiento y evaluación del modelo predictivo
├── docs/
│   └── report_entrega1.pdf                  # Documentación y reporte de entrega
├── logs/
│   └── dolar_analysis_20250508.log          # Registros y logs de ejecución
├── setup.py                               # Archivo de configuración e instalación
└── README.md                             # Documentación principal del proyecto
```

## Automatización con GitHub Actions

Se implementó un flujo de trabajo en `.github/workflows/update_data.yml` que se activa automáticamente cada vez que se realiza un push a la rama main. Este workflow configura el entorno, instala las dependencias necesarias y ejecuta el script principal (`main.py`), encargado de recolectar, transformar y guardar los datos históricos. En caso de detectar nuevos datos, el workflow realiza automáticamente un commit y push al repositorio, garantizando la trazabilidad y actualización continua del histórico sin necesidad de intervención manual.

### Desarrollo en GitHub Codespaces

Todo el entorno de desarrollo se configura y ejecuta en Codespaces, asegurando portabilidad y facilidad de configuración. Se incluyen extensiones recomendadas como:

- Python
- Pylance
- Jupyter
- CSV Editor
- GitHub Copilot

## Resultados

- Se construyó una herramienta funcional y automática para descargar, limpiar y almacenar los datos del dólar australiano.
- Se generó el archivo `historical.csv` con más de 6000 registros desde el año 2000 hasta la fecha actual.
- Se garantiza la persistencia del histórico, ya que las nuevas ejecuciones agregan solo datos no existentes, evitando duplicados.
- Los logs permiten rastrear todas las ejecuciones y errores ocurridos.
- El flujo automatizado de GitHub Actions verifica, ejecuta y versiona cada actualización de manera transparente.
- Todo el proyecto es portable y se ejecuta íntegramente en la nube sin configuración adicional local.

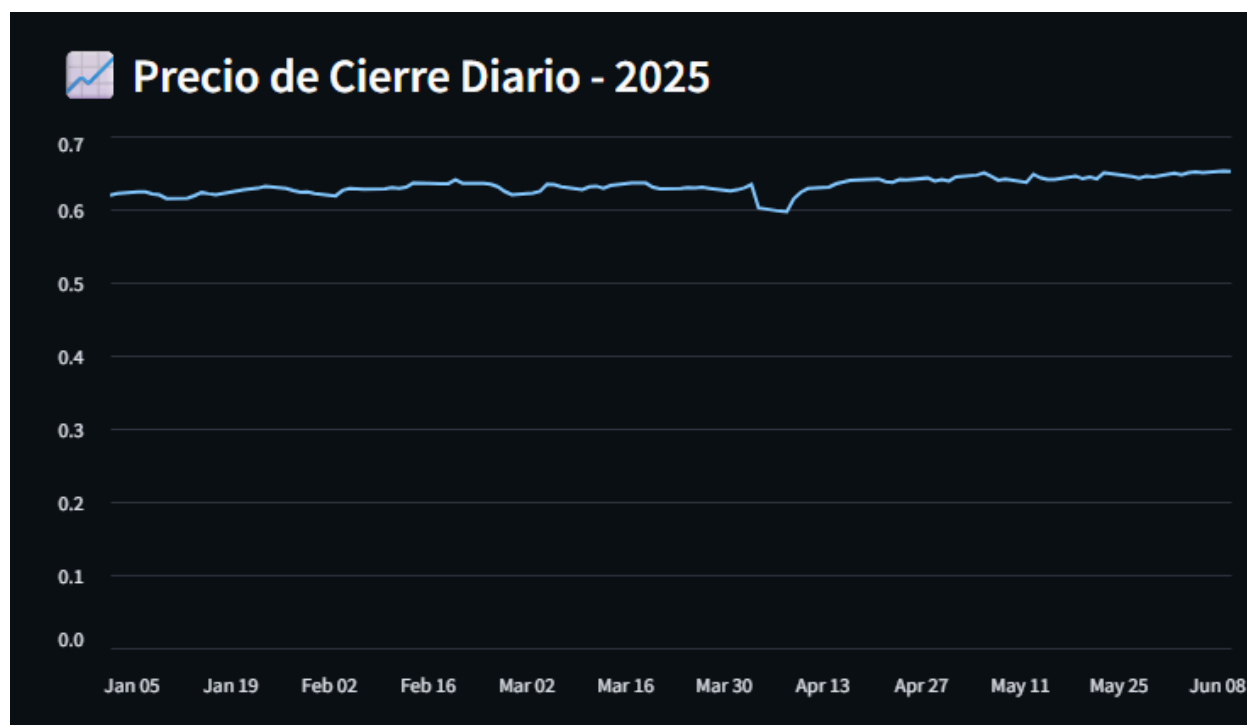
## Reporte final

- Kpis y Métricas

### 1- Evolución del Precio de Cierre Diario

- Qué mide: El valor del contrato de futuros del Dólar Australiano al final de cada día de negociación. Es la métrica base para todos los demás cálculos.
- Métrica (Fórmula): Precio\_Cierre ('Close')

Gráfico:



Interpretación:

- Este gráfico se enfoca en el primer semestre de 2025. El precio del "Aussie" muestra una relativa estabilidad, fluctuando principalmente en un rango entre 0.64 y 0.67.

- Se observa un período de baja volatilidad hasta finales de marzo, seguido de una caída notable a principios de abril y una posterior recuperación. Esto sugiere un evento específico que afectó al mercado en esa fecha.
- La tendencia general en este corto período es lateral, sin una dirección alcista o bajista clara y sostenida.

## 2- Tendencia de Mediano Plazo

- Qué Mide: Compara el precio diario con su tendencia suavizada de los últimos 30 días para identificar la dirección del mercado a mediano plazo y visualizar los principales ciclos económicos.
- Métrica (Fórmula):  $\text{Media Móvil (MA)} = \frac{\text{Suma(Precios de Cierre de los últimos 30 días)}}{30}$

Gráfico:



Interpretación:



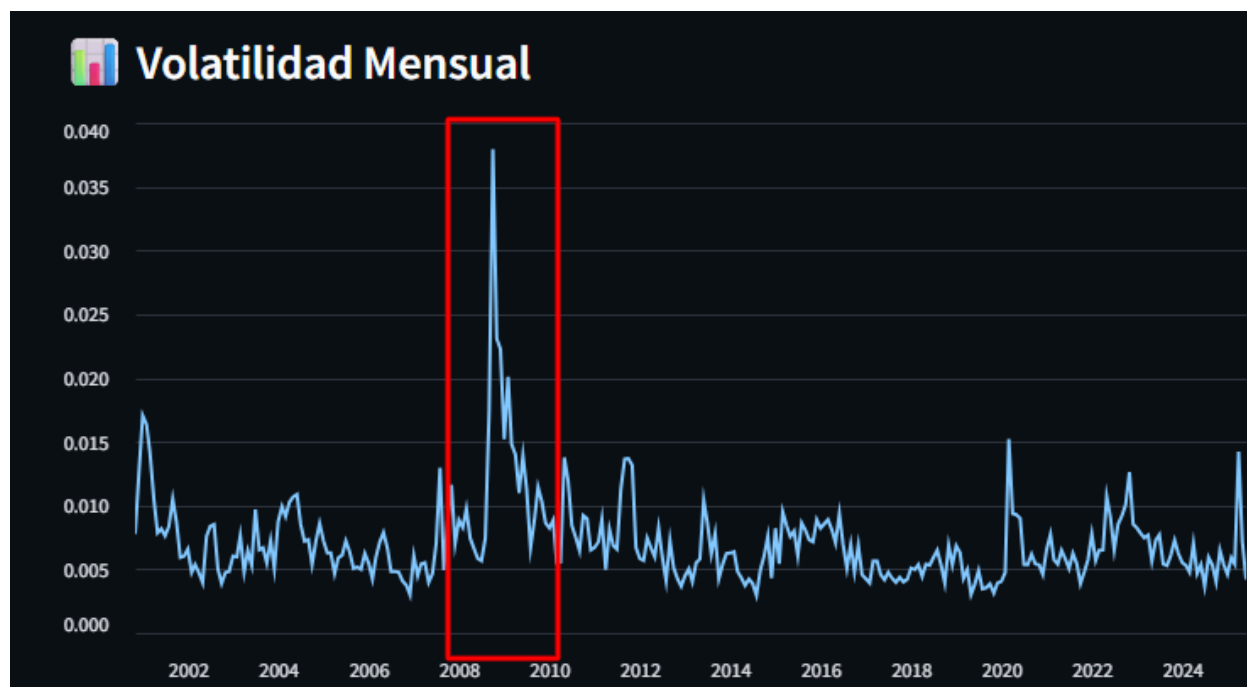
- El gráfico muestra claramente los grandes ciclos económicos. Se observa una fuerte tendencia alcista desde el 2009 hasta el pico de 2011, coincidiendo con el boom de las materias primas post-crisis.
- Desde 2011 hasta aproximadamente 2016 (y de nuevo desde 2021), el precio se ha mantenido mayormente por debajo de su media móvil, indicando prolongados períodos bajistas o de debilidad para el Dólar Australiano.
- El V-Shape de 2020 es evidente, mostrando la drástica caída y rápida recuperación durante la pandemia

### 3- Volatilidad Histórica Mensual

- Qué mide: El grado de dispersión o variación del precio en un mes. Una volatilidad alta significa grandes fluctuaciones de precio (más riesgo y oportunidad); una baja significa estabilidad.
- Métrica / Fórmula:

Desviación Estándar ( $\sigma$ ) de los precios de cierre diarios dentro de un mes

Gráfico:



Interpretación:

- Los inversionistas retiraron capital de mercados emergentes o de riesgo y se refugiaron en activos más seguros como el dólar estadounidense, lo que causó una caída abrupta en el AUD y una gran volatilidad.
- Durante octubre, el AUD cayó fuertemente frente al USD, lo que explica esos picos en el KPI de volatilidad mensual.
- El dato más impactante es el enorme pico de volatilidad en 2008-2009. Esto es la firma inconfundible de la Crisis Financiera Global, un período de pánico extremo en los mercados.
- Se observa un segundo pico, más pequeño pero significativo, en 2020, correspondiente al inicio de la pandemia de COVID-19.
- Fuera de estos eventos de crisis, la volatilidad se ha mantenido en niveles mucho más bajos y controlados, especialmente entre 2016 y 2019.

#### 4- Rendimiento Acumulado de la Inversión (Buy & Hold)

- Que mide: El rendimiento total de una inversión hipotética realizada al inicio del período analizado (año 2000), asumiendo que el activo se mantiene ("hold"). Muestra la rentabilidad a largo plazo.
- Métrica (Fórmula): Producto Acumulado de  $(1 + \text{Retorno Diario})$

Gráfico:



Interpretación:

- Un inversor que compró en el año 2000 habría visto su inversión más que duplicarse hasta el pico de 2011.
- Sin embargo, desde 2011, el rendimiento ha caído drásticamente, mostrando que el Dólar Australiano ha perdido gran parte del valor ganado en la década anterior.

- Actualmente, el retorno acumulado sigue siendo positivo, pero muy por debajo de sus máximos históricos, lo que demuestra que no ha sido una buena inversión "comprar y mantener" en la última década.

### 5- Identificación de Eventos de Crisis

- Qué Mide: Pinpoints las fechas exactas de mayor pánico o euforia en el mercado. Es una herramienta para conectar la volatilidad con eventos históricos específicos.
- Métrica (Fórmula): Conteo de días donde  $|(Cierre\_Hoy - Cierre\_Ayer) / Cierre\_Ayer| > 0.05$

Gráfico:

⚠ Días con Variaciones Extremas (±5%)			
Total de días extremos: 8			
	fecha	cerrar	retorno
4196	2008-10-03 00:00:00		0.7724
4194	2008-10-07 00:00:00		0.7152
4192	2008-10-09 00:00:00		0.6984
4182	2008-10-23 00:00:00		0.6518
4163	2008-11-19 00:00:00		0.6436
4139	2008-12-24 00:00:00		0.6777
4138	2008-12-26 00:00:00		0.7198
46	2025-04-03 00:00:00		0.6338

Interpretación:

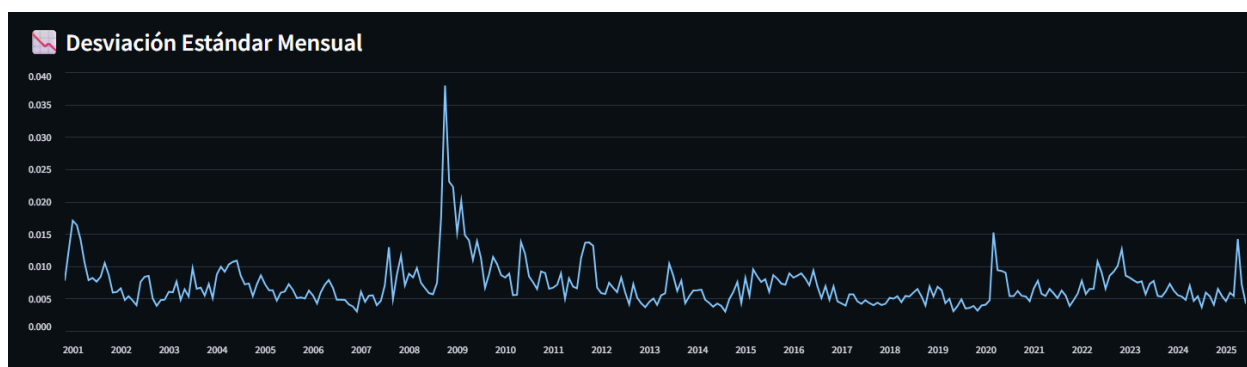
- Se han identificado un total de 8 días de movimientos extremos.
- Todas las fechas listadas pertenecen al último trimestre de 2008. Esto confirma de manera precisa que el período de máximo estrés y pánico del mercado en las últimas dos décadas se concentró durante la fase más aguda de la Crisis Financiera Global, corroborando lo visto en el gráfico de volatilidad.

## 6- Dispersión de Precios Mensual

- Qué mide: La dispersión estadística de los precios de cierre diarios con respecto a su precio promedio durante cada mes. Es la medida cuantitativa más directa de la volatilidad; un valor alto indica que los precios fluctuaron salvajemente, mientras que un valor bajo indica que se mantuvieron estables y predecibles.
- Métrica (Fórmula):  $\sigma = \sqrt{[\Sigma(x_i - \mu)^2 / N]}$

Donde  $x_i$  es cada precio de cierre diario del mes,  $\mu$  es el precio promedio de ese mes, y  $N$  es el número de días de negociación en el mes.

Gráfico:

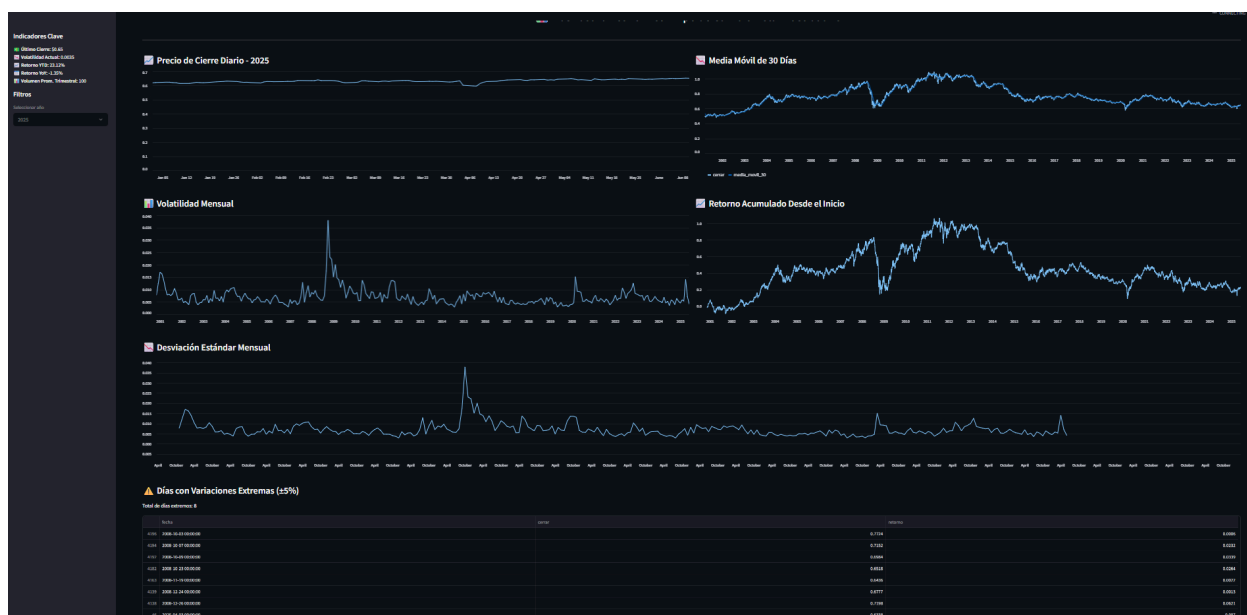


Interpretación:

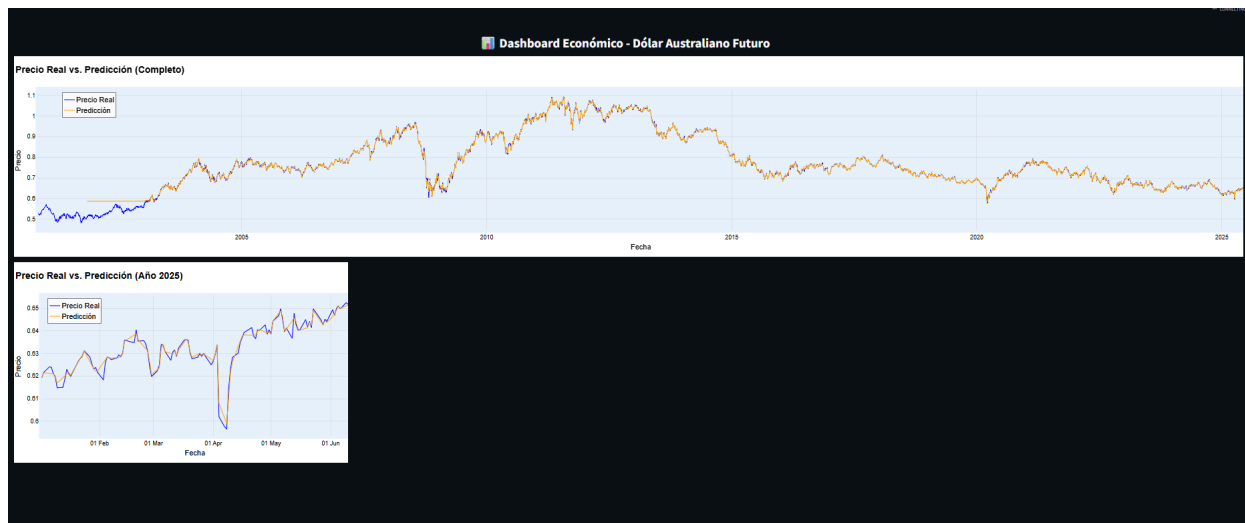
- Este gráfico confirma visualmente los hallazgos del gráfico de "Volatilidad Mensual". El pico más pronunciado y evidente se sitúa en el período 2008-2009, lo que refleja la extrema incertidumbre y las violentas oscilaciones de precios durante la Crisis Financiera Global.
- Se identifica un segundo pico notable en 2020, asociado al shock inicial de la pandemia de COVID-19.
- Los períodos como 2017-2019 muestran una desviación estándar muy baja, lo que indica un mercado con movimientos de precios muy acotados y una volatilidad reducida.

- En esencia, este gráfico funciona como el "sismógrafo" del mercado: registra con precisión la magnitud de los "terremotos" financieros.

## Visualización Dashboard 1.



## Visualización Dashboard 2



### Interpretación:

#### 1. Interpretación del Gráfico: "Precio Real vs. Predicción"

- La línea de predicción (naranja) traza la trayectoria general del precio real (azul) con una fidelidad visual impresionante a lo largo de más de 20 años. Sigue las grandes tendencias alcistas (como la de 2009-2011) y las bajistas (como la de 2011-2016) de manera muy cercana.

### Análisis Crítico:

- La predicción sigue al precio real demasiado bien. Los mercados financieros son inherentemente "ruidosos" y difíciles de predecir con tanta exactitud.
- El comportamiento más probable que se observa aquí es el de un modelo "ingenuo" o "naive". Este tipo de modelo realiza una predicción muy simple: "el precio de mañana será el precio de hoy".

Por ejemplo, es como conducir un carro mirando exclusivamente por el espejo retrovisor. Sabes exactamente dónde has estado, pero no tienes idea de hacia dónde vas. Un modelo así siempre parecerá correcto un día después, pero no ofrece ninguna ventaja para anticipar el siguiente movimiento.

## 2. Interpretación del Gráfico: "Precio Real vs. Predicción (Año 2025)"

- El zoom en 2025 permite ver con más detalle la alta correlación entre ambas series. La predicción captura los pequeños giros y vueltas del precio real.

Análisis Crítico :

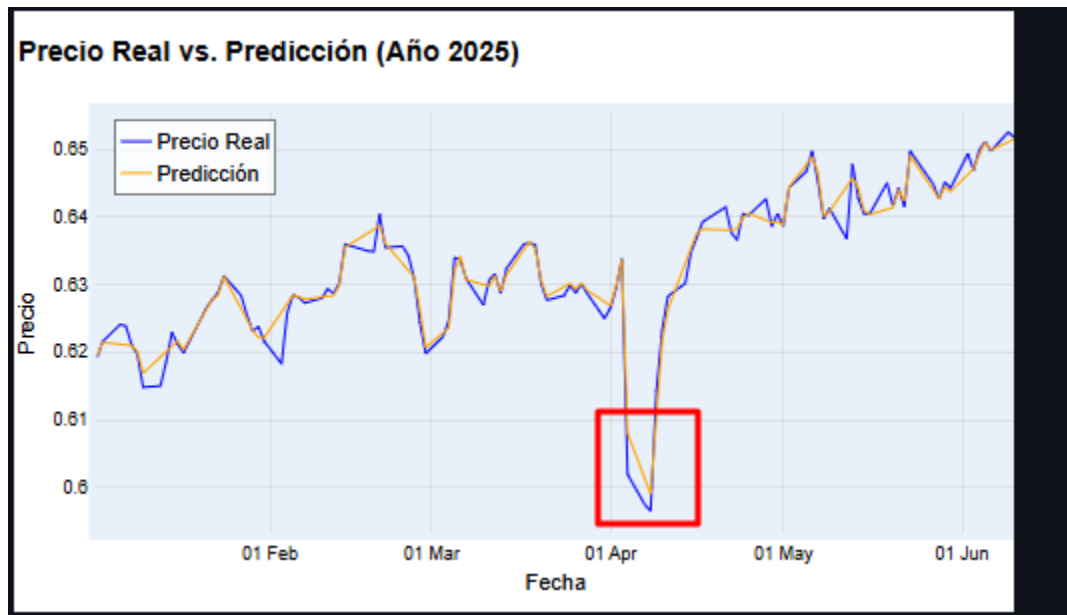
- Este gráfico confirma la sospecha del modelo ingenuo. es de detallar que en los picos y valles agudos. Por ejemplo, en la fuerte caída alrededor del 1 de abril:

El precio real (azul) cae bruscamente.

La predicción (naranja) replica esa misma caída, pero un paso (un día) después.

El modelo no anticipa la caída; simplemente reacciona a ella una vez que ha ocurrido. Esto se repite en casi todos los movimientos bruscos. La línea naranja es consistentemente un "eco" de la línea azul.





**Link de la presentación:**

<https://gamma.app/docs/Australian-Dollar-Futures-6AF-duhw50gpplcm8yo>

## Bibliografía

- GitHub. (2024). *GitHub Actions documentation*. <https://docs.github.com/actions>
- Python Software Foundation. (2024). *Python language reference*.  
<https://docs.python.org/3/>
- Scikit-learn developers. (2024). *Scikit-learn: Machine learning in Python*.  
<https://scikit-learn.org/>
- Statsmodels developers. (2024). *Statsmodels: Statistical modeling in Python*.  
<https://www.statsmodels.org/stable/>
- Yahoo Finanzas. (2024). *Historial de cotizaciones de AUD/USD*.  
<https://es.finance.yahoo.com/quote/6A%3DF/history/>
- Git Development Community. (2024). Git (versión 2.40) [Software]. Recuperado de  
<https://git-scm.com/>