

**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 Codespaces. 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, versionado de código y despliegue continuo, lo que facilita su escalabilidad y adaptación a otros activos financieros similares.

## **Objetivo General**

Comprobar la habilidad para 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

- Precio promedio mensual/anual de cierre

Para analizar la evolución del tipo de cambio AUD/USD en diferentes periodos.

- Volatilidad histórica mensual

Calculada como la desviación estándar del precio de cierre.

Indica qué tan variable ha sido el mercado en un periodo específico.

- Tendencia interanual (YoY)

Comparación del cierre anual con respecto al año anterior.

Identifica tendencias de apreciación o depreciación del AUD.

- Máximo y mínimo histórico

Puntos extremos del precio desde el año 2000.

Permite establecer umbrales de soporte o resistencia técnica.

- Volumen promedio de negociación por trimestre

Muestra la actividad del mercado y el interés en este futuro.

- Días con cambios extremos (+/-5%)

Identificación de eventos macroeconómicos relevantes.



## 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/
│       │   │   └── historical.csv    # Dataset generado
│       │   └── models/
│       ├── collector.py             # Clase principal de recolección
│       ├── logger.py               # Configuración de logs
│       └── main.py                 # Punto de entrada del programa
├── docs/
│   └── report_entrega1.pdf          # Documentación
├── logs/
│   └── dolar_analysis_20250508.log  # Registros de ejecución
├── setup.py
└── README.md
```

## 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.

## 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/>