

Evidencia de Aprendizaje 1

Proyecto Integrado V - Línea de Énfasis (Entrega 1)

Estudiantes:

Maria Johana Arenas Usuga

Docente: Andrés Felipe Callejas

Curso: Proyecto Integrado V – PREICA2501B020128

Universidad Digital de Antioquia

Ingeniería de Software y Datos

Mayo de 2025

Automatización del Análisis de Datos Financieros de Bancolombia S.A. (CIB) utilizando Python y GitHub Actions

Resumen

Este informe describe el desarrollo de un sistema automatizado para la recolección y análisis de datos históricos de la acción de Bancolombia S.A. (CIB), disponible públicamente en la plataforma Yahoo Finance. El proyecto fue desarrollado en Python y se compone de módulos independientes para la recolección de datos (collector), enriquecimiento con indicadores financieros (enricher), modelado (modeller) y registro de eventos (logger). La automatización del flujo de trabajo se logró mediante la integración de GitHub Actions, lo que permite programar la actualización y análisis de los datos de forma continua y sin intervención manual. El propósito de esta solución es facilitar la obtención y procesamiento de información financiera actualizada que sirva de base para la toma de decisiones estratégicas o académicas en el contexto de inteligencia de negocios.

Introducción

El análisis de datos financieros es una práctica esencial para inversionistas, analistas y entidades interesadas en entender el comportamiento del mercado. Con la disponibilidad de información bursátil en línea, se han abierto nuevas oportunidades

para automatizar el acceso y tratamiento de estos datos. En este contexto, la acción de Bancolombia S.A. (CIB), listada en la Bolsa de Nueva York (NYSE), fue seleccionada como caso de estudio por su relevancia en el sistema financiero latinoamericano.

Yahoo Finance, una de las plataformas más utilizadas para consultar información bursátil, proporciona datos históricos de precios accesibles públicamente. Aprovechando esta disponibilidad, se diseñó un sistema automatizado que extrae dicha información y la transforma en un conjunto de datos enriquecido con indicadores clave como volatilidad y media móvil. Además, el sistema está pensado para operar de forma recurrente, sin necesidad de intervención humana, gracias al uso de flujos de trabajo automatizados con GitHub Actions.

Metodología

La solución fue desarrollada bajo una arquitectura modular, donde cada componente del proceso cumple una función específica, permitiendo escalabilidad y mantenibilidad:

Extracción de datos (Collector): Se construyó un scraper en Python que utiliza requests y BeautifulSoup para obtener datos históricos de la acción CIB desde Yahoo Finance. La información se procesa

y se convierte en un DataFrame de pandas, eliminando filas vacías o irrelevantes, y luego se almacena en formato CSV.

Enriquecimiento de datos (Enricher): Sobre la base de datos recolectada, se calculan indicadores financieros como volatilidad (desviación estándar móvil), retornos, y otros KPI financieros que permiten analizar el comportamiento del activo en el tiempo.

Registro de eventos (Logger): Se implementó un sistema de logging personalizado que registra cada evento relevante del proceso, desde la extracción hasta el guardado de datos, en archivos de texto ubicados en una carpeta logs/.

Automatización del flujo de trabajo: Mediante la configuración de GitHub Actions (`.github/workflows/update_data.yml`), se programó la ejecución automática del script `main.py` para asegurar actualizaciones periódicas del dataset.

Ejecución del sistema: El archivo principal `main.py` orquesta el flujo completo, integrando los módulos anteriores. La ejecución se realiza desde la terminal con el comando `python src/maria_piv/main.py`.

Conclusión

El desarrollo de este sistema demuestra cómo las herramientas de código abierto y los servicios en la nube pueden combinarse para automatizar procesos de análisis financiero. La integración de Python con GitHub Actions permite crear soluciones reproducibles, programables y mantenibles, que reducen la dependencia de la intervención humana en tareas repetitivas. Este enfoque no solo mejora la eficiencia operativa, sino que también proporciona una base robusta para el análisis de datos financieros en tiempo casi real. La solución presentada puede adaptarse fácilmente a otras acciones o indicadores económicos, ampliando su utilidad tanto para fines académicos como empresariales.