

PROYECTO INTEGRADO V - LINEA DE ÉNFASIS (ENTREGA 1)

Automatización de la recolección de datos históricos financieros del Grupo Aval mediante Github Actions y Python

Estudiante:

Jean Carlos Páez Ramírez (Grupo: Preica2501b020128)

Docente:

Andrés Felipe Callejas Jaramillo

Especialista en Analítica de Datos

Curso:

Proyecto Integrado V - Línea de Énfasis

Programa de Ingeniería en Desarrollo de Software y Datos

Facultad Ingeniería y Ciencias Agropecuarias

Institución Universitaria Digital de Antioquia

Contenido

Resumen	3
Introducción.....	4
Metodología.....	5
Bibliografía	7

Resumen

Este informe describe el desarrollo de un sistema automatizado para la recolección, almacenamiento y trazabilidad de datos históricos del Grupo Aval, una de las entidades financieras más representativas de Colombia. El proyecto emplea herramientas de programación en Python, específicamente las librerías `yfinance` y `pandas`, para descargar información desde Yahoo Finanzas. Se incorpora un enfoque de programación orientada a objetos (OOP), junto con un sistema de registros (logging) estructurado en formato CSV. Este sistema fue implementado mediante un módulo especializado (`csv_logger.py`), que centraliza la trazabilidad en el archivo `log_data.csv`, con el fin de mantener un control preciso de cada ejecución del proceso. La automatización se logra mediante la plataforma GitHub Actions, que permite actualizar diariamente los datos financieros a las 12:00 UTC. Esta solución proporciona una base sólida para el análisis económico de activos bursátiles en tiempo casi real, facilitando procesos de toma de decisiones, modelado de datos y evaluación del comportamiento financiero de la entidad.

Palabras clave: Grupo Aval, automatización, GitHub Actions, `yfinance`, Python, recolección de datos históricos, análisis económico.

Introducción

El análisis de series de tiempo financieras constituye un elemento fundamental en la toma de decisiones dentro del ámbito económico y empresarial. En este contexto, contar con datos históricos confiables, actualizados y fácilmente accesibles es esencial para elaborar modelos predictivos y realizar estudios de comportamiento de activos bursátiles. Este proyecto se centra en el Grupo Aval Acciones y Valores S.A. (AVAL), cuyo comportamiento en el mercado de valores resulta representativo del sistema financiero colombiano.

La automatización de procesos mejora la eficiencia de tareas repetitivas y propensas a errores, permitiendo a las áreas de negocio enfocarse en actividades de mayor valor agregado (Danta Analytics, s.f.).

El presente trabajo describe el diseño e implementación de un sistema automatizado para la recolección diaria de información financiera histórica de AVAL. Para ello, se emplean herramientas de desarrollo en Python y procesos de integración continua mediante GitHub Actions, lo que garantiza la actualización diaria de los datos y la trazabilidad de cada ejecución. Esta solución, enfocada en el aprendizaje práctico de automatización, forma parte de la asignatura “Proyecto Integrado V”, dentro de la línea de énfasis en análisis económico.

Metodología

El desarrollo del proyecto se llevó a cabo siguiendo una metodología basada en la implementación iterativa de componentes funcionales, estructurada de la siguiente manera:

Selección del activo financiero: Se eligió el Grupo Aval (símbolo bursátil AVAL) como objeto de análisis, dada su relevancia en el mercado financiero colombiano. Los datos históricos de AVAL se obtuvieron desde Yahoo Finanzas, una fuente confiable y ampliamente utilizada para información financiera (Yahoo Finance, s.f.).

Desarrollo del colector de datos: Se implementó un script en Python utilizando las librerías `yfinance` y `pandas`. Este script obtiene datos históricos de AVAL desde Yahoo Finanzas y los almacena en un archivo `historical.csv`. La biblioteca `pandas` proporciona estructuras de datos de alto rendimiento y herramientas de análisis fáciles de usar para el lenguaje de programación Python (Reback et al., 2020).

Módulo de logging personalizado: Para garantizar la trazabilidad del proceso, se desarrolló un sistema de registros basado en la librería `logging`, con un manejador personalizado implementado en el módulo `csv_logger.py`. Este módulo permite registrar de forma estructurada los eventos relevantes de cada ejecución en un archivo centralizado (`log_data.csv`) ubicado en el directorio `src/logs`. El registro en formato CSV complementa los archivos de texto `.log`, brindando un historial tabular que facilita el análisis automatizado, la auditoría y el seguimiento del funcionamiento del sistema (Python Software Foundation, 2020).

Automatización con GitHub Actions: Se configuró un flujo de trabajo (`update_data.yml`) que ejecuta el script diariamente a las 12:00 UTC. Este flujo permite la descarga automática de nuevos datos, la actualización del archivo histórico y el registro de la ejecución en los logs. GitHub

Actions permite automatizar, personalizar y ejecutar flujos de trabajo de desarrollo de software directamente en el repositorio de GitHub (GitHub, s.f.).

Estructura del proyecto: El repositorio se organizó con una estructura modular clara y documentada, incluyendo el archivo `setup.py` para facilitar la instalación del proyecto como paquete.

Esta metodología permitió el diseño de una solución escalable y fácilmente integrable en proyectos de análisis financiero o académico.

Bibliografía

Danta Analytics. (s.f.). APA - Automatización de rutinas de análisis. Recuperado de <https://dantaanalytics.com/servicios/apa/>

GitHub. (s.f.). GitHub Actions documentation. Recuperado de <https://docs.github.com/actions>

Python Software Foundation. (2020). logging — Logging facility for Python. Recuperado de <https://docs.python.org/3/library/logging.html>

Reback, J., McKinney, W., & Wes McKinney. (2020). pandas: powerful Python data analysis toolkit. Recuperado de <https://pandas.pydata.org/>

Yahoo Finance. (s.f.). Yahoo Finance - AVAL. Recuperado de <https://es-us.finanzas.yahoo.com/quote/AVAL/>