



EVIDENCIA DE APRENDIZAJE 1

ANÁLISIS Y HERRAMIENTAS DE EXTRACCIÓN DE DATOS

JULIANA MARÍA PEÑA SUAREZ

PROGRAMACIÓN PARA ANÁLISIS DE DATOS - PREICA2501B020065

ANDRÉS FELIPE CALLEJAS JARAMILLO

MEDELLIN-2025

CONTENIDO

INTRODUCCION	3
DESCRIPCION DE LA PAGINA Y ARTICULO A ANALIZAR	4
DESCRIPCION DEL TEMA DE INTERES	5
OBJETIVOS	6
Objetivo General:.....	6
Objetivos Específicos:	6
METODOLOGIA EMPLEADA EN EL SCRAPING	7
Estructura del Proyecto:.....	7
Flujo General del Scraper	8
Herramientas y librerías utilizadas.....	8
RESULTADOS Y CONCLUSIONES	9
BIBLIOGRAFIA	10

INTRODUCCION

El análisis de datos económicos es fundamental para comprender las dinámicas del mercado y tomar decisiones informadas. En este contexto, la automatización de la recolección de datos mediante técnicas de web scraping se ha convertido en una herramienta esencial para empresas que dependen de información financiera precisa y oportuna.

Este proyecto fue desarrollado en **Medellín, Antioquia**, durante el año **2025**, como parte de una práctica académica aplicada al entorno local. La recolección automatizada de datos históricos del petróleo crudo WTI (CL=F), extraídos desde el portal Yahoo Finance, busca simular una solución tecnológica orientada a **Energéticos del Norte S.A.S.**, una empresa del sector energético interesada en integrar datos económicos a sus análisis de proyecciones de costos, tendencias de mercado y toma de decisiones estratégicas.

El presente informe describe detalladamente el proceso de scraping utilizando la biblioteca **BeautifulSoup**, así como la organización del código, los resultados obtenidos y las posibles aplicaciones del sistema desarrollado.

DESCRIPCION DE LA PAGINA Y ARTICULO A ANALIZAR

Yahoo Finance es una plataforma reconocida que proporciona información financiera en tiempo real, incluyendo precios de acciones, noticias económicas y datos históricos de diversos activos. La sección seleccionada para este proyecto es la de datos históricos del petróleo WTI:

<https://es.finance.yahoo.com/quote/CL=F/history/>

Esta página presenta una tabla con los precios diarios de apertura, cierre, máximos, mínimos y volumen de negociación del petróleo WTI, información crucial para Análisis económicos y financieros.

DESCRIPCION DEL TEMA DE INTERES

El petróleo WTI es un referente clave en los mercados energéticos globales. Su precio influye en diversas áreas, desde la inflación hasta las políticas monetarias. Analizar su comportamiento histórico permite identificar tendencias, evaluar riesgos y desarrollar modelos predictivos. Este proyecto se centra en automatizar la extracción de estos datos para facilitar su análisis posterior.

OBJETIVOS

Objetivo General:

Desarrollar una herramienta automatizada que permita extraer y almacenar datos históricos del petróleo WTI desde Yahoo Finance utilizando para el scraping utilizando la biblioteca **BeautifulSoup**,

Objetivos Específicos:

- Comprender la estructura HTML de la página de Yahoo Finance para identificar los elementos relevantes.
- Implementar un script en Python que utilice BeautifulSoup para navegar y extraer los datos deseados.
- Almacenar los datos extraídos en un formato estructurado (CSV) para su análisis posterior.
- Evaluar la eficiencia y robustez de la herramienta desarrollada.

METODOLOGIA EMPLEADA EN EL SCRAPING

La metodología empleada en este proyecto se fundamenta en la implementación de un proceso automatizado de recolección de datos desde el sitio [Yahoo Finance](#), con apoyo en herramientas del ecosistema de Python. El enfoque fue modular, basado en buenas prácticas de ingeniería de software, lo cual facilita la mantenibilidad y escalabilidad del sistema.

El desarrollo se encuentra documentado en el repositorio de GitHub:

 https://github.com/Julianapenaestudiante/pad_2025_1_2

Estructura del Proyecto:

1. **dataweb.py**

Contiene la clase DataWeb, que realiza las siguientes funciones:

- Envía una solicitud HTTP a la URL de Yahoo Finance correspondiente al histórico del petróleo WTI (CL=F).
- Utiliza BeautifulSoup para analizar el HTML y localizar la tabla de datos históricos.
- Extrae las filas de la tabla y organiza los datos en un DataFrame de pandas.

2. **main.py**

Este archivo actúa como punto de entrada del programa:

- Importa la clase DataWeb desde dataweb.py.
- Instancia un objeto de DataWeb y llama al método obtener_datos().

- Si se obtienen datos, los guarda en un archivo CSV llamado data_web.csv.

Flujo General del Scraper

1. Inicialización del scraper desde main.py.
2. Solicitud HTTP a la página destino con requests.
3. Parsing HTML con BeautifulSoup para localizar la tabla.
4. Extracción de datos estructurados en formato tabular.
5. Transformación y exportación a CSV mediante pandas.

Herramientas y librerías utilizadas

- requests: para acceder al contenido HTML de la web.
- BeautifulSoup: para procesar y extraer elementos HTML.
- pandas: para estructurar los datos y manipularlos fácilmente.
- GitHub: como repositorio de control de versiones y colaboración.

RESULTADOS Y CONCLUSIONES

La herramienta desarrollada permitió automatizar exitosamente la extracción de datos históricos del petróleo WTI desde el portal Yahoo Finance, logrando almacenarlos en un formato estructurado (CSV) apto para análisis posteriores. Este resultado valida la efectividad del uso de bibliotecas del ecosistema Python, como requests, BeautifulSoup y pandas, en procesos de scraping orientados a la recolección de indicadores económicos públicos.

El proyecto, desarrollado en la ciudad de **Medellín, Antioquia, durante el año 2025**, se enmarca dentro de una práctica académica aplicada con proyección profesional. Su implementación fue concebida como una solución tecnológica adaptable a contextos reales, como el de la empresa **Energéticos del Norte S.A.S.**, la cual requiere integrar datos financieros actualizados en sus análisis de costos, tendencias de mercado y toma de decisiones estratégicas.

La arquitectura modular del código facilita su mantenimiento, escalabilidad y adaptación a otros dominios económicos. Además, constituye una base sólida para integrarse en sistemas de analítica avanzada, modelos predictivos y automatización de reportes periódicos, lo que potencia su valor práctico dentro de sectores que dependen de información financiera precisa y oportuna.

BIBLIOGRAFIA

- Scrapy. (2025). Scrapy Documentation. Recuperado de <https://docs.scrapy.org/>
- Yahoo Finance. (2025). CL=F - Crude Oil Futures Historical Data. Recuperado de <https://es.finance.yahoo.com/quote/CL=F/history/>
- Python Software Foundation. (2025). *Python (Versión 3.10) [Software]*. Recuperado de <https://www.python.org/>
- BeautifulSoup. (2025). *BeautifulSoup Documentation*. Recuperado de <https://www.crummy.com/software/BeautifulSoup/>
- Git. (2025). *Git Documentation*. Recuperado de <https://git-scm.com/doc>
- Microsoft. (2025). *Visual Studio Code Documentation*. Recuperado de <https://code.visualstudio.com/doc>
- McKinney, W. (2022). *Python for Data Analysis* (3ra ed.). O'Reilly Media.
- Cedeño, L. (2021). *Minería de datos aplicada con Python*. Alfaomega Grupo Editor.