

Case Study: Global Commodity Price Data Automation (983K+ Records)

Project Overview

This project involved the development and deployment of an automated web scraping system to extract comprehensive global commodity price data from `dailymetalprice.com`. The primary objective was to provide a continuously updated, structured dataset of commodity prices, enabling in-depth market analysis, forecasting, and strategic decision-making for clients in various industries.

The Challenge

Extracting historical and ongoing commodity price data from `dailymetalprice.com` presented several key challenges:

- **High Volume & Historical Depth:** Scraping nearly 1 million records spanning from 2011 to 2025 required an efficient and robust solution.
- **Dynamic Content & Anti-Bot Measures:** The website utilizes JavaScript for content rendering and has mechanisms to deter automated scraping, necessitating advanced browser automation techniques.
- **Currency and Unit Variations:** Prices are available in multiple currencies (USD, GBP, AUD, CNY, BTC, EUR) and units (lb, oz), requiring careful handling and extraction.
- **Automated Scheduling:** The need for a system that could run automatically on a monthly basis to ensure data freshness.
- **Reliability & Data Quality:** Ensuring consistent, accurate, and complete data extraction over extended periods.

The Solution

A sophisticated and resilient web scraping and automation system was engineered using Python and Selenium, deployed on Azure, to address these challenges.

1. Technology Stack

- **Python:** The core programming language for the entire system.
- **Selenium:** Utilized for its powerful browser automation capabilities. Selenium was crucial for navigating the dynamic website, handling JavaScript-rendered content, and interacting with elements like currency dropdowns. It allowed for simulating human browsing behavior to bypass anti-bot measures.
- **Pandas:** Employed for efficient post-extraction data cleaning, transformation, and structuring into analytical-ready DataFrames.
- **Azure:** The system was deployed on Microsoft Azure, leveraging its cloud capabilities for automated, scheduled execution.
- **Excel:** Data was delivered in a clean, structured Excel format, suitable for direct import into databases or analytical tools.

2. Automated Workflow

The system implemented a multi-layered approach to data extraction and automation:

- **Intelligent Navigation & Ad Handling:** The scraper was designed to navigate to the archive pages, handle potential ad interruptions by refreshing, and then systematically iterate through years (2011-2025) and individual date links.
- **Dynamic Content & Currency Selection:** Selenium was used to expand year sections, click on date links, and dynamically change currency selections (USD, GBP, AUD, CNY, BTC, EUR) for each date to capture all available price data.
- **Robust Data Extraction:** Custom logic was developed to reliably extract commodity, price, currency, and unit from the HTML tables, even with variations in price formatting.
- **Automated Scheduling (Azure Deployment):** The entire system was deployed on Azure, configured to run automatically on a monthly basis. This ensures that the client always has access to the most up-to-date commodity price data without manual intervention.

- **Data Cleaning & Consolidation:** Extracted data was processed using Pandas to ensure consistency, handle missing values, and consolidate data from different currencies and dates into a single, comprehensive dataset.
- **Output Generation:** The cleaned and structured data was continuously appended to an Excel file, ensuring a single, growing repository of historical and current commodity prices.

Results

This project successfully delivered a highly effective and automated solution for commodity price data extraction:

- **Vast & Up-to-Date Dataset:** Successfully scraped 983,176 records from 2011 to 2025, providing a rich, continuously updated source of global commodity prices.
- **Automated & Reliable:** The Azure-deployed system runs monthly without manual oversight, ensuring consistent data flow and high reliability.
- **Comprehensive Data:** Captured prices across multiple commodities, currencies, and units, offering a holistic view of the market.
- **Actionable Insights:** The structured dataset empowers clients to perform in-depth market analysis, identify trends, and make informed strategic decisions related to commodity trading and procurement.
- **Scalability & Maintainability:** The modular design allows for easy expansion to include new commodities or adapt to website changes.

Conclusion

This project demonstrates my expertise in developing, deploying, and maintaining large-scale, automated web scraping solutions for critical business intelligence. My proficiency in Python, Selenium, Pandas, and cloud deployment (Azure), combined with a deep understanding of data processing, enables me to deliver high-quality, actionable data that drives significant business value. This experience is directly applicable to any organization seeking reliable, automated data streams for market analysis, financial modeling, or strategic planning.

GitHub Repository

Explore the code and project details on GitHub: <https://github.com/hajrawajid/commodity-price-scraper-python>

Upwork Profile

Connect with me on Upwork: https://www.upwork.com/freelancers/hajrawajid?mp_source=share