

Scraping Product Prices with Python

Objective:

The goal of this project is to develop a web scraping script that extracts pricing and product information from Amazon. Students will use Python and essential third-party libraries to gather and analyze data, helping businesses make informed decisions.

Project Requirements:

1. Setup Environment:

- Ensure Python 3.x is installed.
- Install necessary libraries using:
`pip install requests beautifulsoup4`

2. Understanding Web Scraping:

- Learn the basics of HTTP requests and responses.
- Understand how to inspect web pages using browser developer tools.

3. Scraping Amazon Data:

- Use the `requests` library to fetch HTML content from an Amazon product page.
- Utilize `BeautifulSoup` to parse and extract relevant details, such as:
 - Product Name
 - Price
 - Ratings & Reviews
 - Availability

4. Data Storage & Output:

- Store the extracted data in a structured format (CSV/JSON file or a database).
- Print the extracted data in a user-friendly format.

5. Handling Anti-Scraping Measures:

- Implement request headers to mimic human browsing.
- Add delays between requests to prevent blocking.
- Discuss ethical scraping and Amazon's `robots.txt` policy.

6. Enhancement & Future Scope:

- Extend the script to track price changes over time.
 - Automate notifications for price drops.
 - Use visualization tools to analyze pricing trends.
-

Deliverables:

- Python script with proper documentation and comments.
- A report explaining the methodology, challenges, and future improvements.
- Sample extracted data (CSV or JSON file).

Evaluation Criteria:

- Code correctness and efficiency.
 - Proper usage of `requests` and `BeautifulSoup`.
 - Ability to extract and store meaningful data.
 - Handling of anti-scraping mechanisms.
 - Documentation and presentation of the findings.
-

Submission Deadline: 06-03-2025

Happy Coding!

Reference:- <https://oxylabs.io/blog/how-to-scrape-amazon-product-data>