

Amazon Scraper

Overview

The script automates the process of scraping product details from Amazon's best-seller category pages. It collects information like product name, price, ratings, seller details, description, and available images, and saves the data to a CSV file.

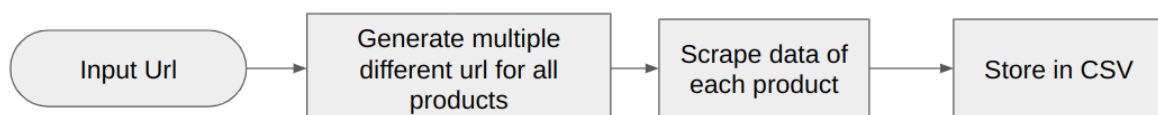
ETL Pipeline:

The script can be described as a basic ETL pipeline, performing the following stages:

1. **Extract:** Fetches data from Amazon's best-seller pages, including product details and images.
2. **Transform:** Processes and organizes the extracted data into a structured format.
3. **Load:** Saves the structured data into a CSV file for further use or analysis.

Key Features

- Extracts product links from multiple Amazon best-seller categories.
- Scrapes detailed information for each product.
- Saves the scraped data into a structured CSV file.



Setup Instructions

Install Required Libraries

Ensure the following Python libraries are installed:

1. pip install selenium
2. Download WebDriver
 - Download the ChromeDriver matching your Google Chrome version from [ChromeDriver Downloads](#).
 - Add the ChromeDriver executable to your system's PATH.
3. Save the Script
 - Save the script to a Python file, amazon.py.

Usage Instructions

1. Update the Bestseller URLs
 - a. Modify the `bestseller_urls` list in the script to include the Amazon category URLs you want to scrape.
2. Run the Script Execute the script using
 - a. `python3 amazon_scraper.py`
3. Output

The scraped product details will be saved to a CSV file named `amazon_products.csv` in the script's directory.

CSV File Details

The generated CSV file contains the following columns:

- Product Name: Name of the product.
- Product Price: Price of the product.
- Best Seller Rating: Best seller badge rating.
- Ship From: Shipping origin.
- Sold By: Seller name.
- Product Rating: Customer ratings.
- Product Description: Key features of the product.
- Number Bought in the Past Month: Purchase stats.
- Category Name: Product category.
- All Available Images: List of image URLs.

*I did not add the sale discount because I was unsure whether it referred to the discount percentage or the amount deducted from the original price.