**What is Selenium?**

* An open-source framework mainly used for **automating web browsers**.
* Primary purpose: **Software Testing** (e.g., simulate clicks, form submissions, navigation).
* Written originally in **Java**, with bindings available for **Python**, **C#**, etc.

Dekho jo basic primary use hota hai Selenium ka

wo hota hai for software testing yaani ki aap log

python-mei command issue krne we tell selenium ki chrome mei kis element pe click krna hai

kya click krne pe kya cheez aegi

tk maani jaayegi in sab cheezon ke liye

Selenium ka istemal kiya jaata hai aaj bhi mukhya roop me.

**2. Why use Selenium?**

* Allows issuing commands to a real browser (e.g., Chrome) to mimic user actions.
* Still widely used for automation tasks in testing and scraping.

Is video me hum log focus karenge

Selenium ko kis tarah se as a web scraping tool use karna hai.

Ab kyon aap use karna chahoge Selenium ko as a web scraping tool?

Iska ek primary reason yeh hai ki jab aap Selenium ka istemal karte ho

toh actual me aapka chrome browser khulta hai ,and we can open chrome and see activity happening!

**Selenium for Web Scraping**

* Websites often **block scraping** done via requests or simple scripts.
* Selenium, however, **mimics real user behavior**, making it harder to detect and block.
* Advantage: You can **see the browser in action** while data is being scraped.

Normally websites scraping ko block kar deti hain, request module ko aur bhi kisi tarah ki automated scaping ko

kyunki wo nahi chahte ki aap log unka data uthao.

Lekin agar aap log Selenium ka istemal karte ho toh wo kaafi close hota hai ek actual user ke behavior ke, aur aap logon ko zyada se zyada data mil sakta hai.

**Proxies with Selenium**

* Proxies help hide your IP and avoid detection/blocks while scraping.
* Example service: **Webshare Proxy**.
* Benefits of using proxies:
  + Prevents rate-limiting or bans.
  + Enables accessing geo-restricted content.
  + Makes scraping appear more like natural traffic.

Iske baad hum log use karenge proxy ka

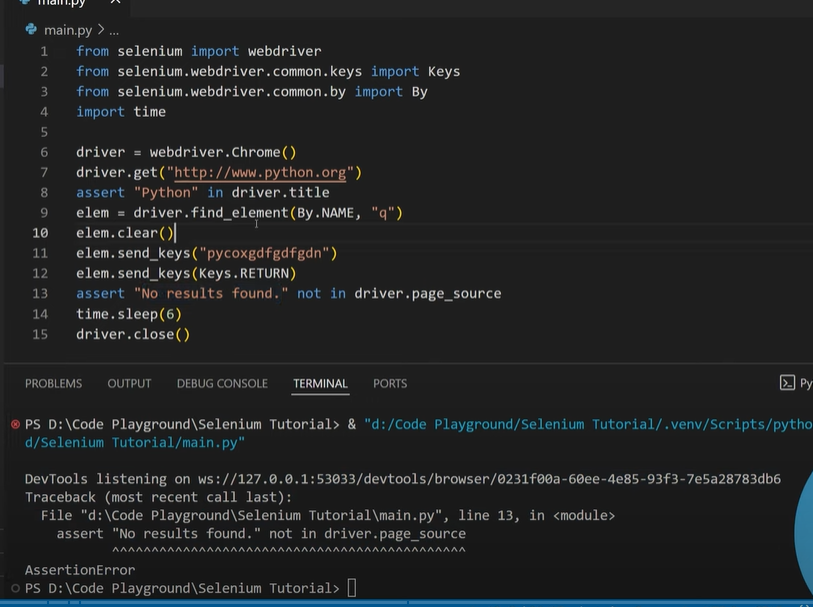
aur proxy ko main loonga Webshare se

aur main aap logon ko dikhaunga ki Webshare ki proxy

kyon aap logon ko istemal karni chahiye

aur kya fayda hai Webshare ki proxy ko use karne me.

Webshare is...



Agar assertion error aya to testing apki fail ho gayi wrna success

Keys.RETURN likhenge to ENTER key chali jaegi

Webshare is proxy provider

**📌 Proxy Server – Concept**

**1. Basic Idea**

* In normal client-server communication:
  + **Client → Server** (direct request)
  + Server receives:
    - Your **IP address**
    - Browser details
    - Other basic information (data points)
* Using this, the server can **identify and track** you.

**2. Problem**

* If you send **too many requests** to a website (e.g., *amazononline.in*),
  + The server may detect unusual activity.
  + Risk: **Block / Ban** due to high request frequency.

**3. Proxy Server Role**

* Acts as a **middleman** between **Client and Server**.
* Flow:
  + Client → Proxy Server → Target Server
  + Target Server → Proxy Server → Client
* So, the **real server** only sees the **proxy’s IP**, not yours.

**4. Benefits**

* Hides **real IP address**.
* Prevents **detection / blocking** during heavy requests.
* Allows distribution of traffic across **multiple proxies** for safe large-scale scraping.

Like mom is ps , and she talks to dad to get ur wish fulfilled

Download proxies and we set manually in our setting ,

Rotating proxies automatically changes

Proj:amazon ke search results ko scape karenge

1st :hwo to locate single and multiple elements

We open amazon.in in gues mode:search alptop and copy url

Plan:locate multiple elements and unko locate krke for loop ke andar outerHTML ko save kar lunga and later I will write a script jo in data ko scrape karegi

Why?

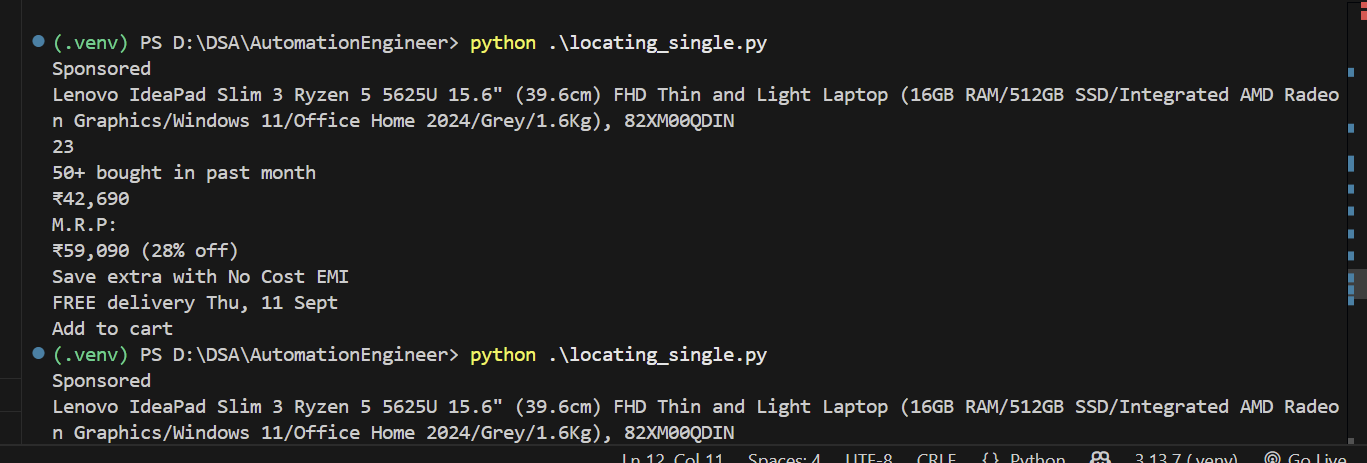
When doing web scraping best practice is to : HTML pages ko scrape kar lo

Challenging part in web scraping: data ko lekar ana apne computer mei

Ex:1million product pages of amazon, 1st we download all those pages and then do 1by1 scraping, becoz on the fly info ko extract karenge to prob ye hogi ki info miss kardo

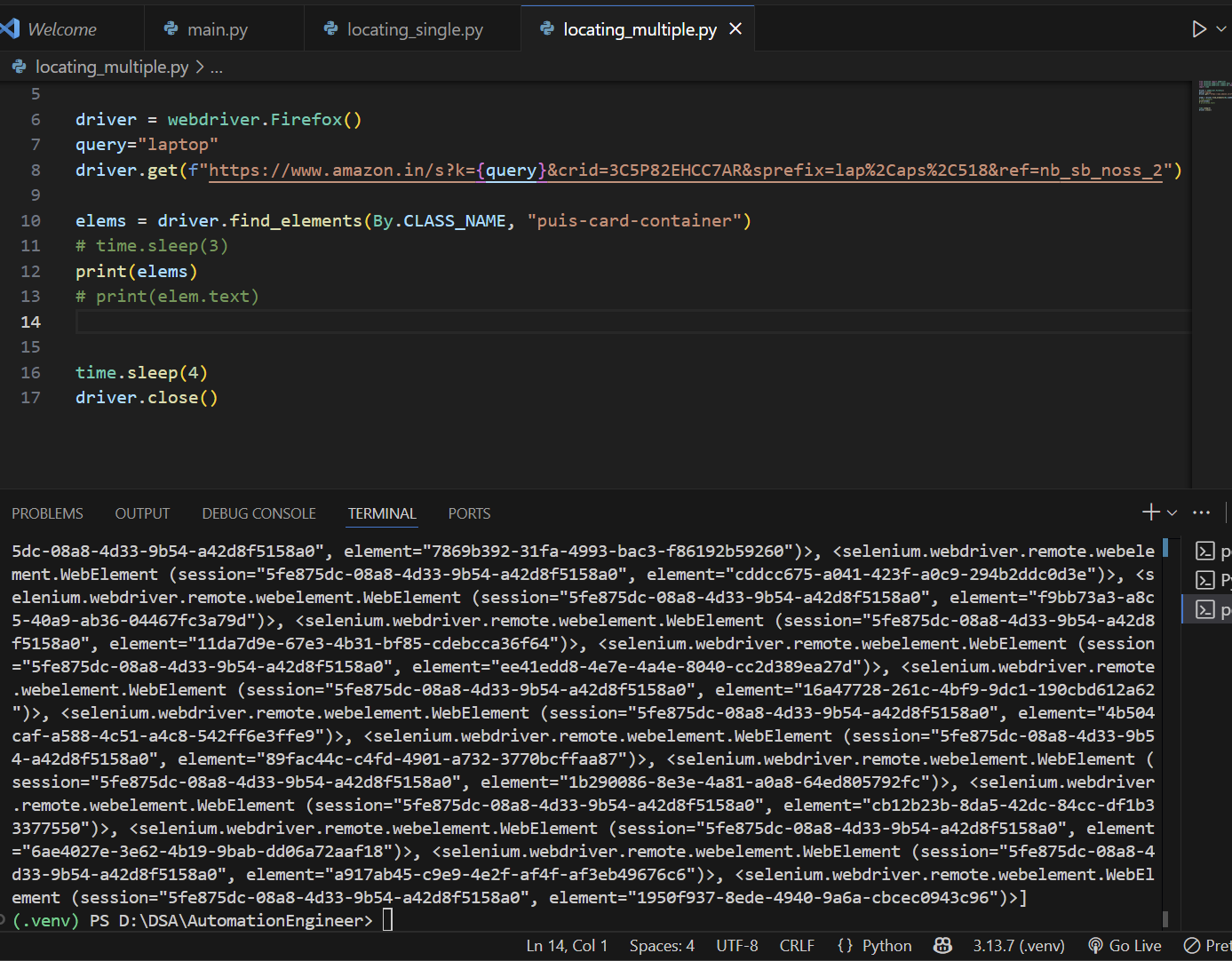
;;;;;;;

Now locating\_multiple.py : kaise multiple data ko locate kar sko



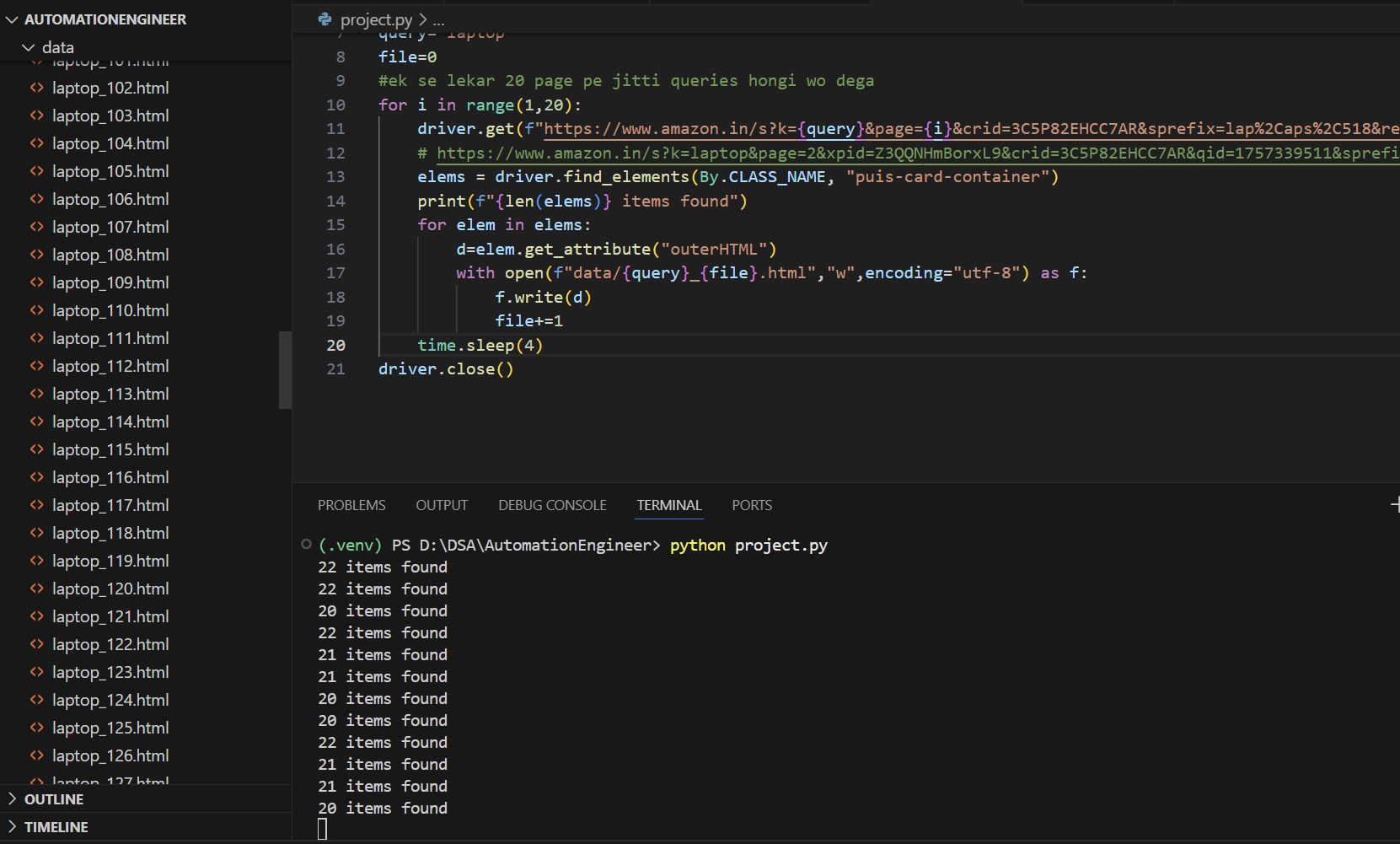
A screen shot of a computer code

AI-generated content may be incorrect.

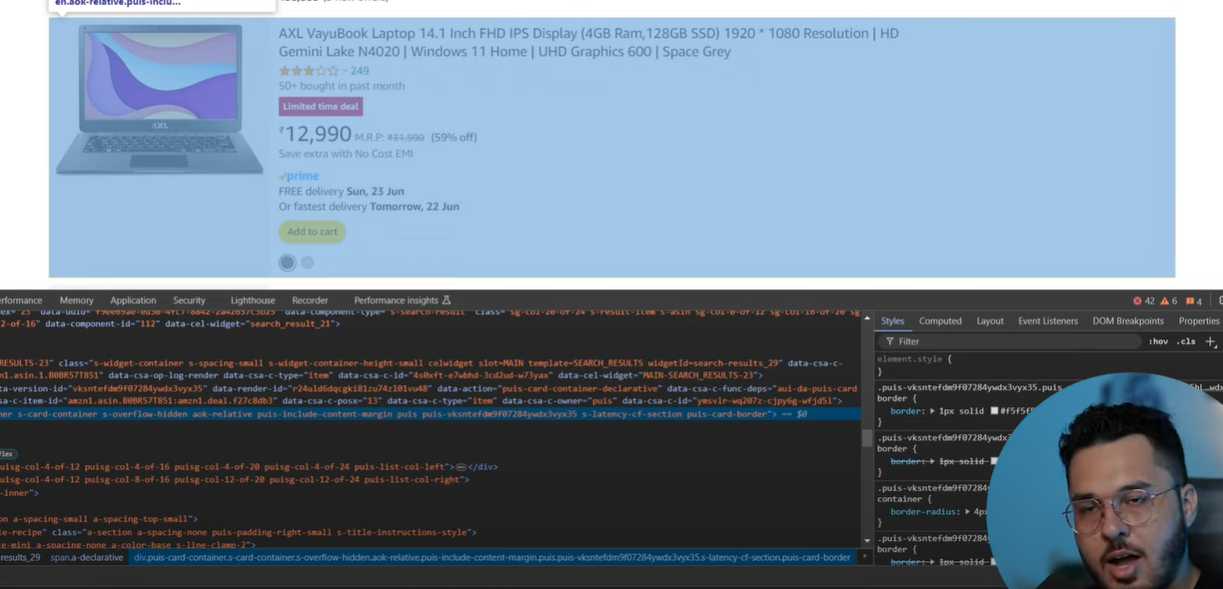
in above we get a list

A screenshot of a computer program

AI-generated content may be incorrect.\

now jab hamare paas aa gaya itna sara data to hum likhenge collection script

Rn ,we wrote query:laptop and ek ke karke saare product ke cards ke html ko save kar lenge

this highlishted blue is a card

Now we have our data in html format, so we use collection script to help collect data

We wont use selenium but we will use beautiful soup

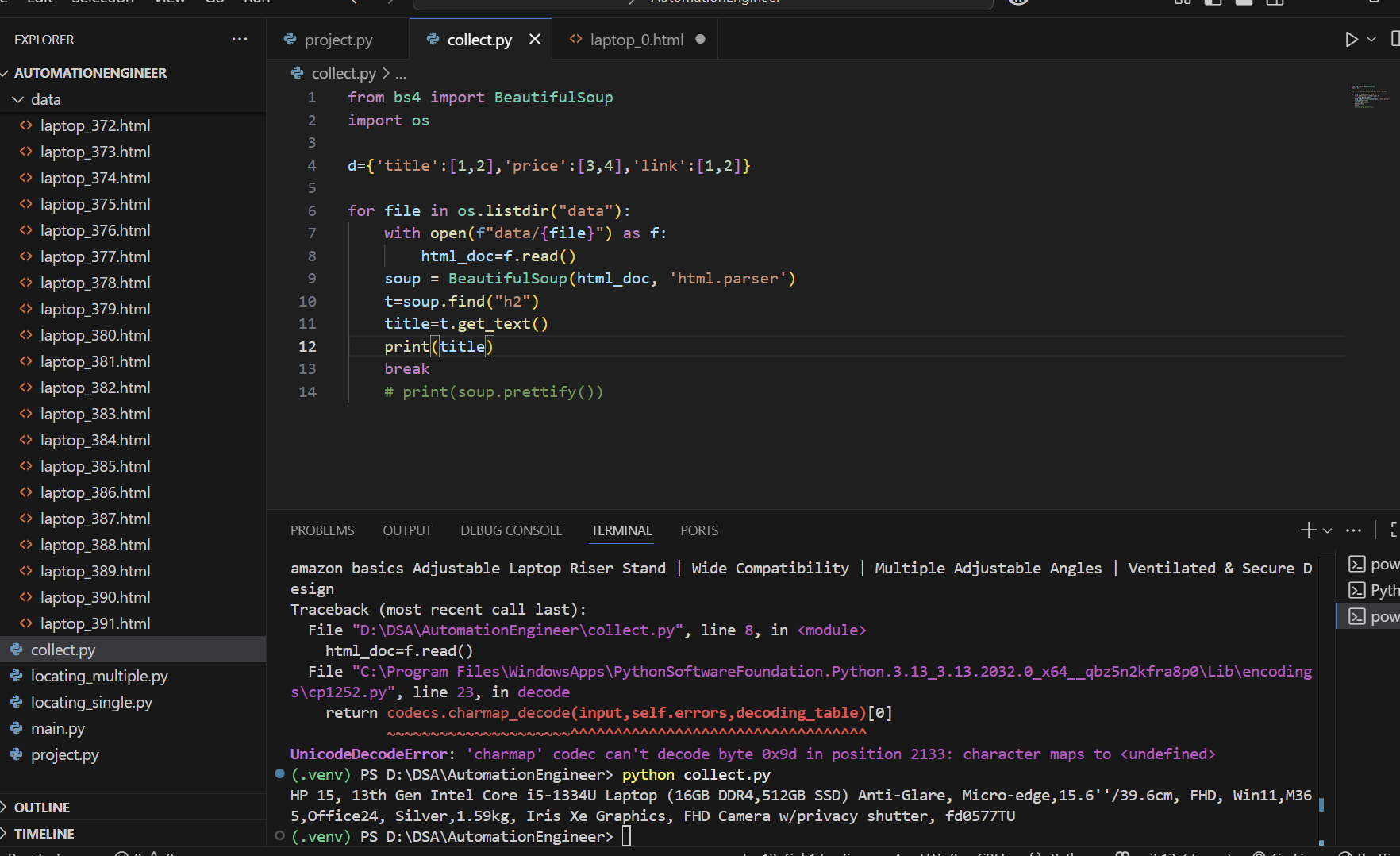
a **much better approach** is to combine **Selenium + BeautifulSoup**:

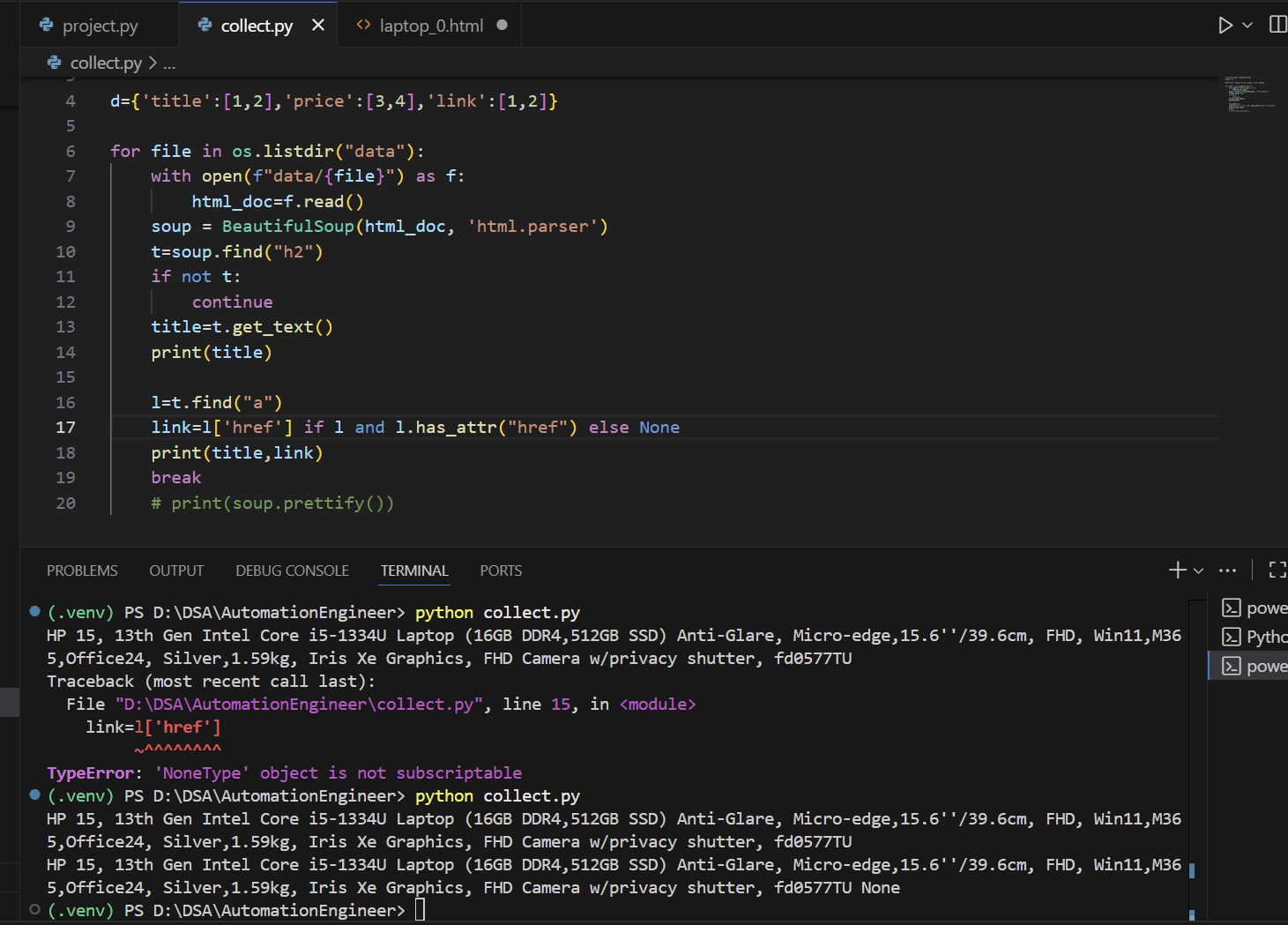
* **Selenium** → loads the dynamic Amazon page (since Amazon uses JavaScript).
* **BeautifulSoup** → parses the HTML and extracts structured data (title, price, rating, etc.) instead of raw HTML blobs.

Pip install bs4

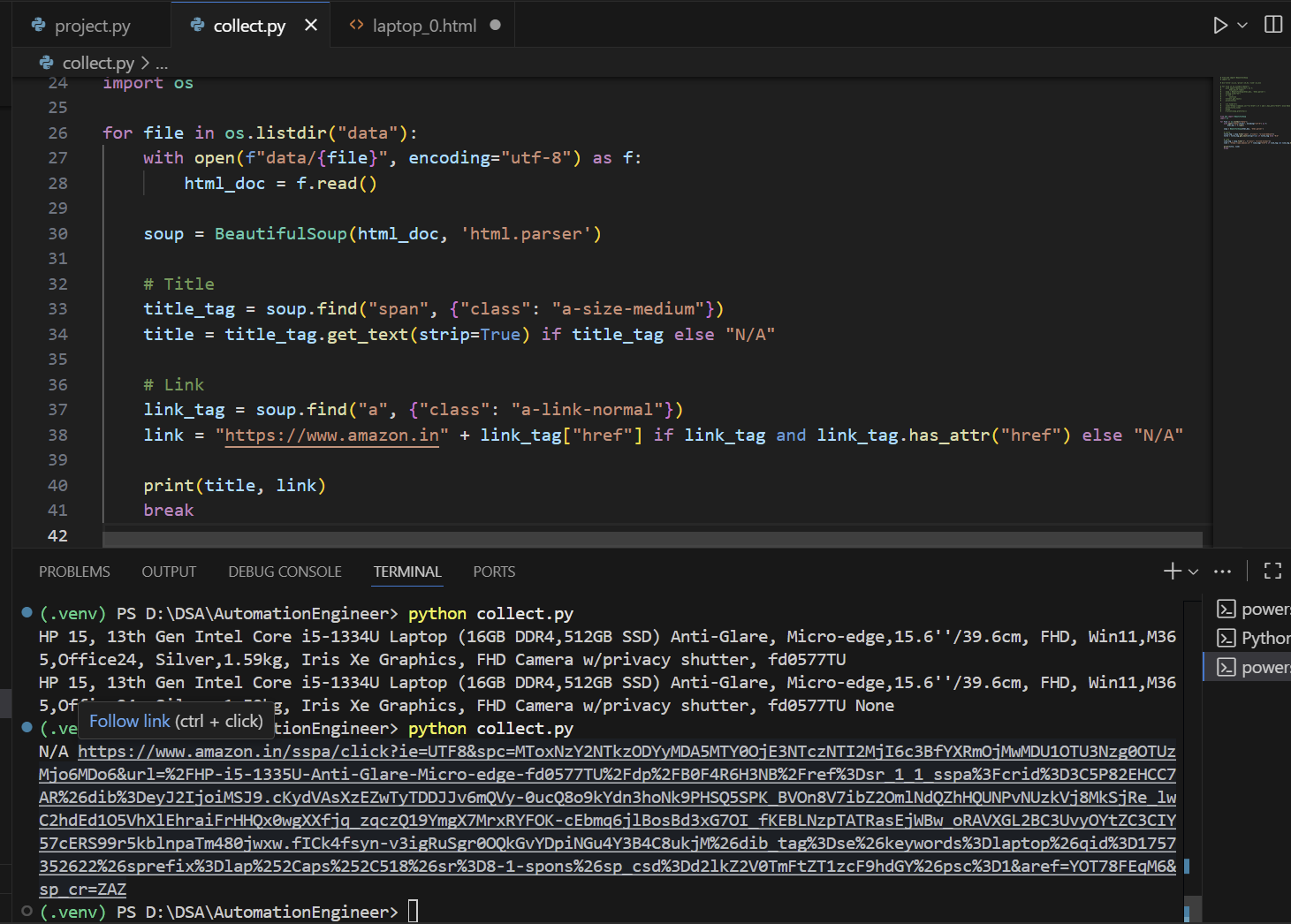
Pip install selenium

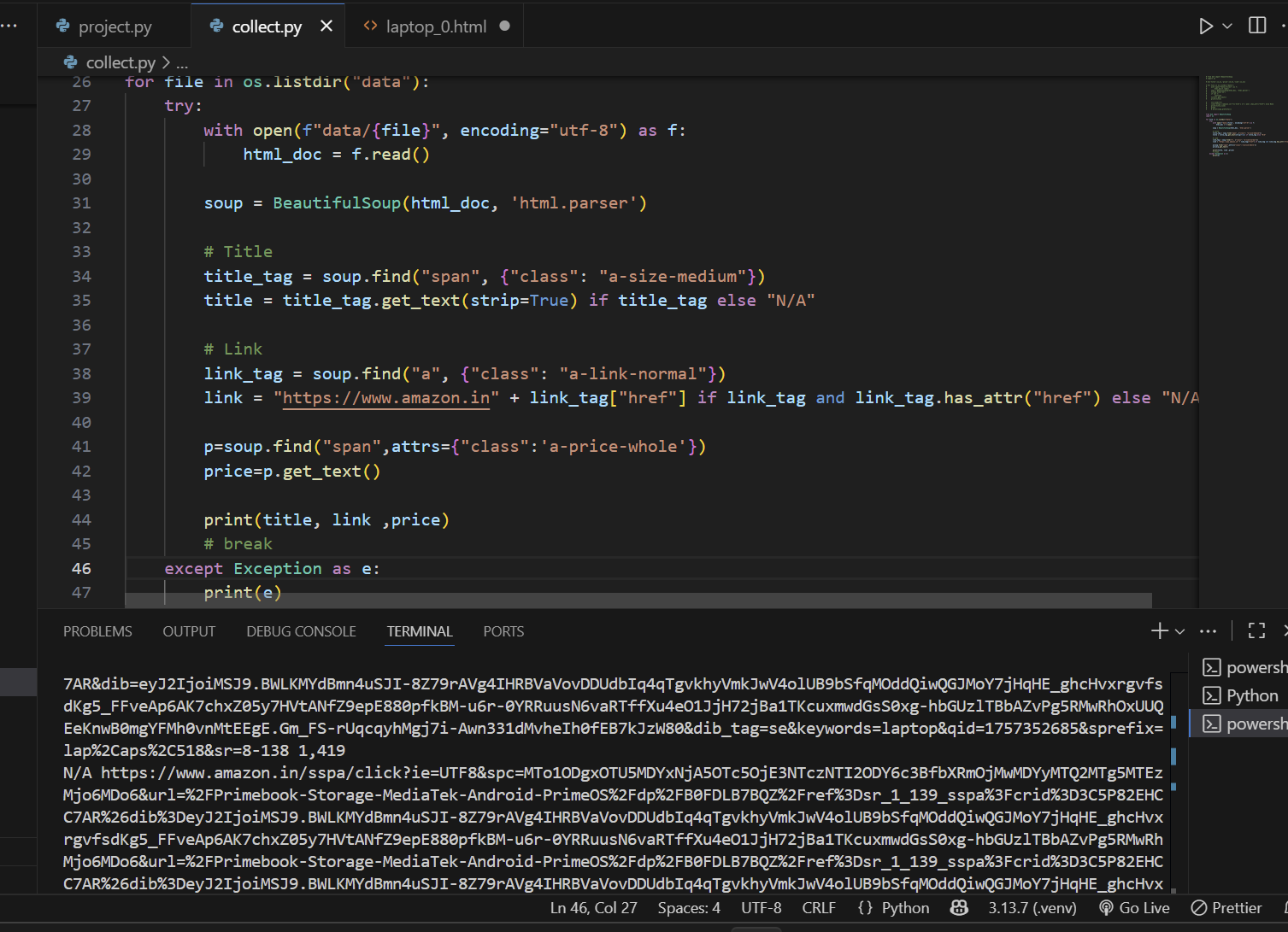
Csv write karenge pandas ke through

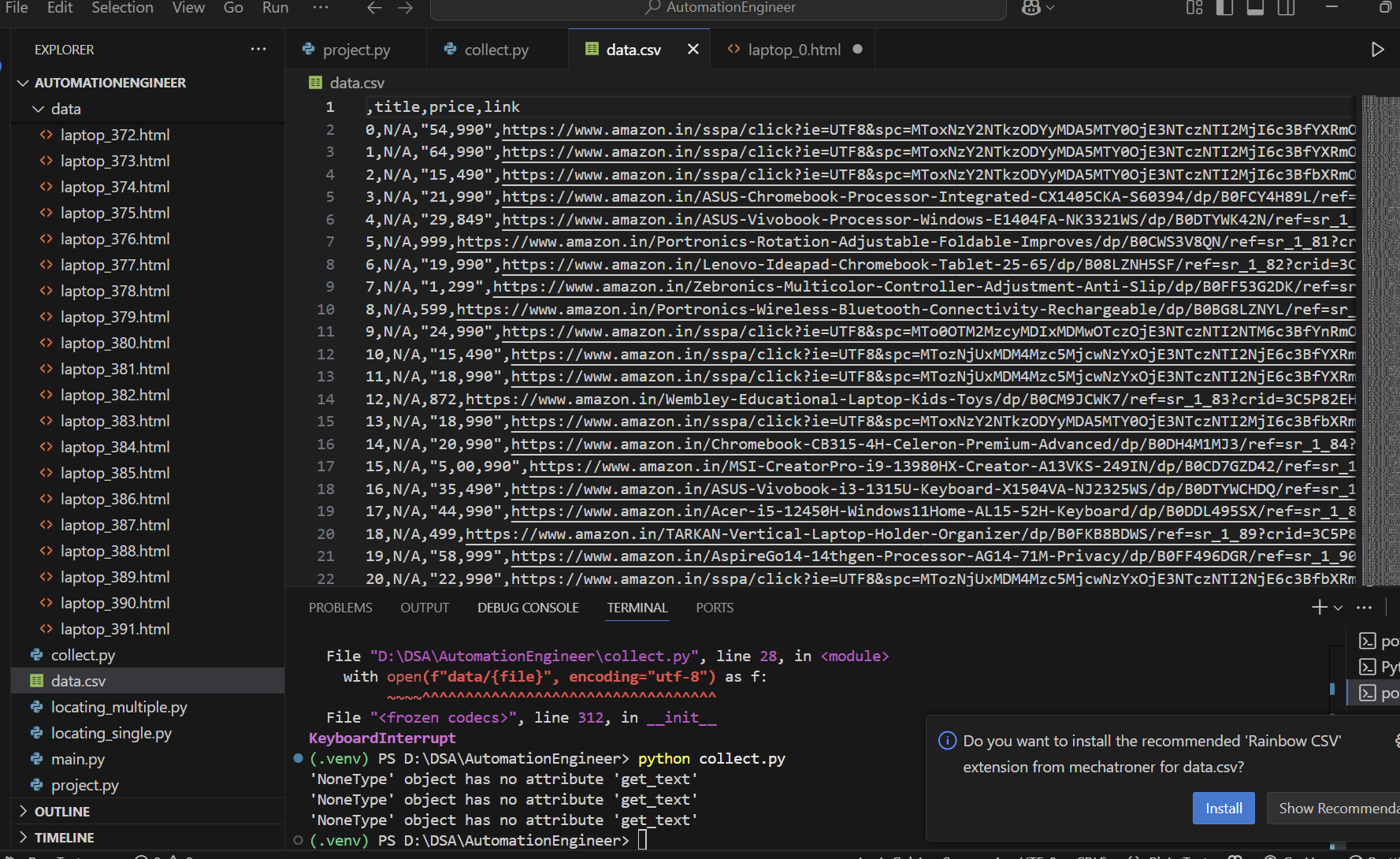




For link:







A screenshot of a computer program

AI-generated content may be incorrect.