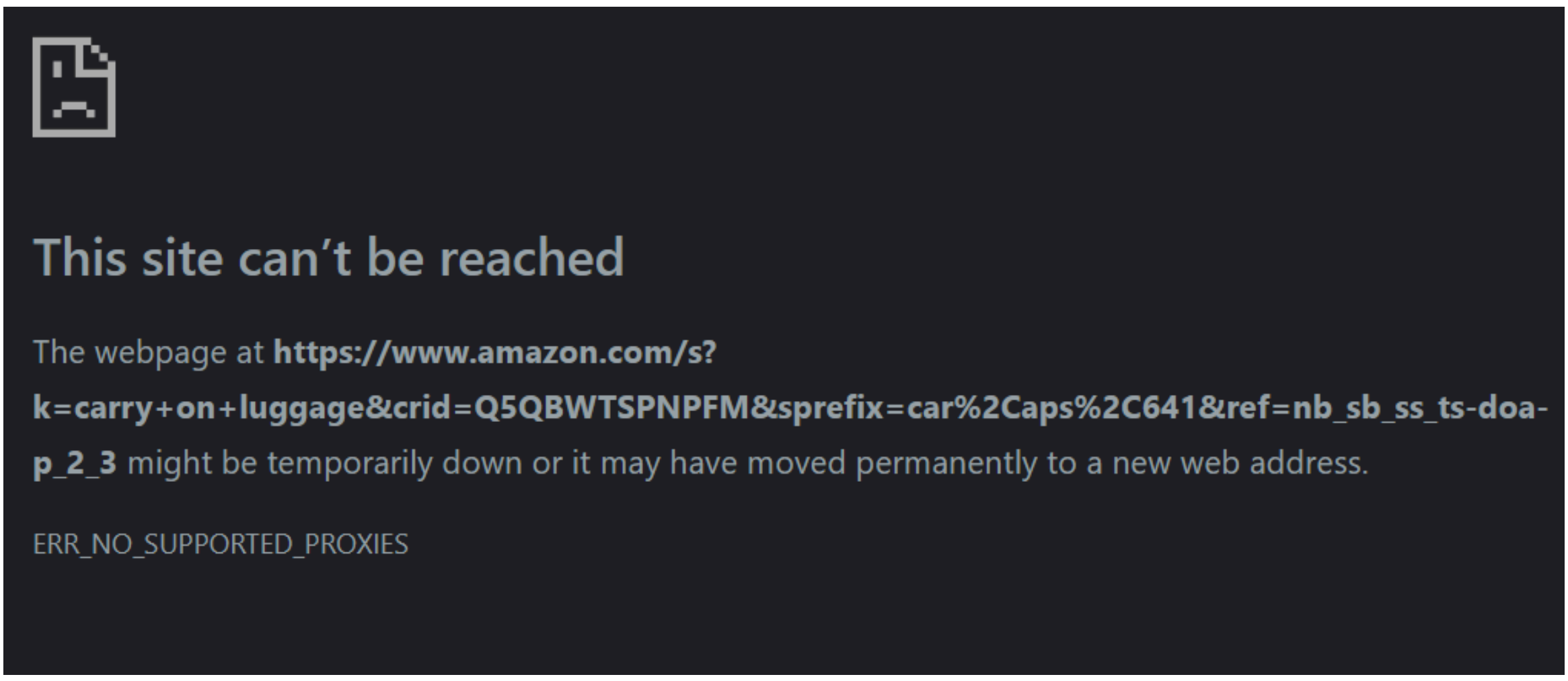
Scraping Amazon.com using Python

Challenges:

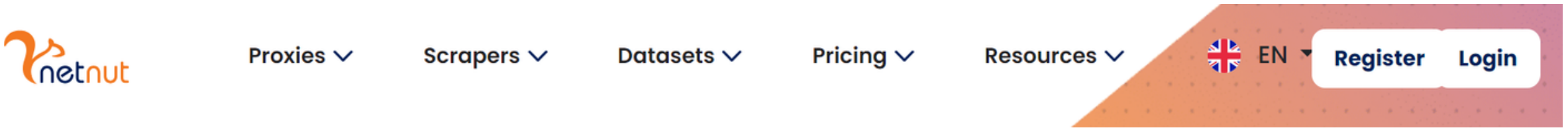
- Avoiding Captcha checks
- Fetching as much data as needed

Trying to Access Amazon.com without Proxy



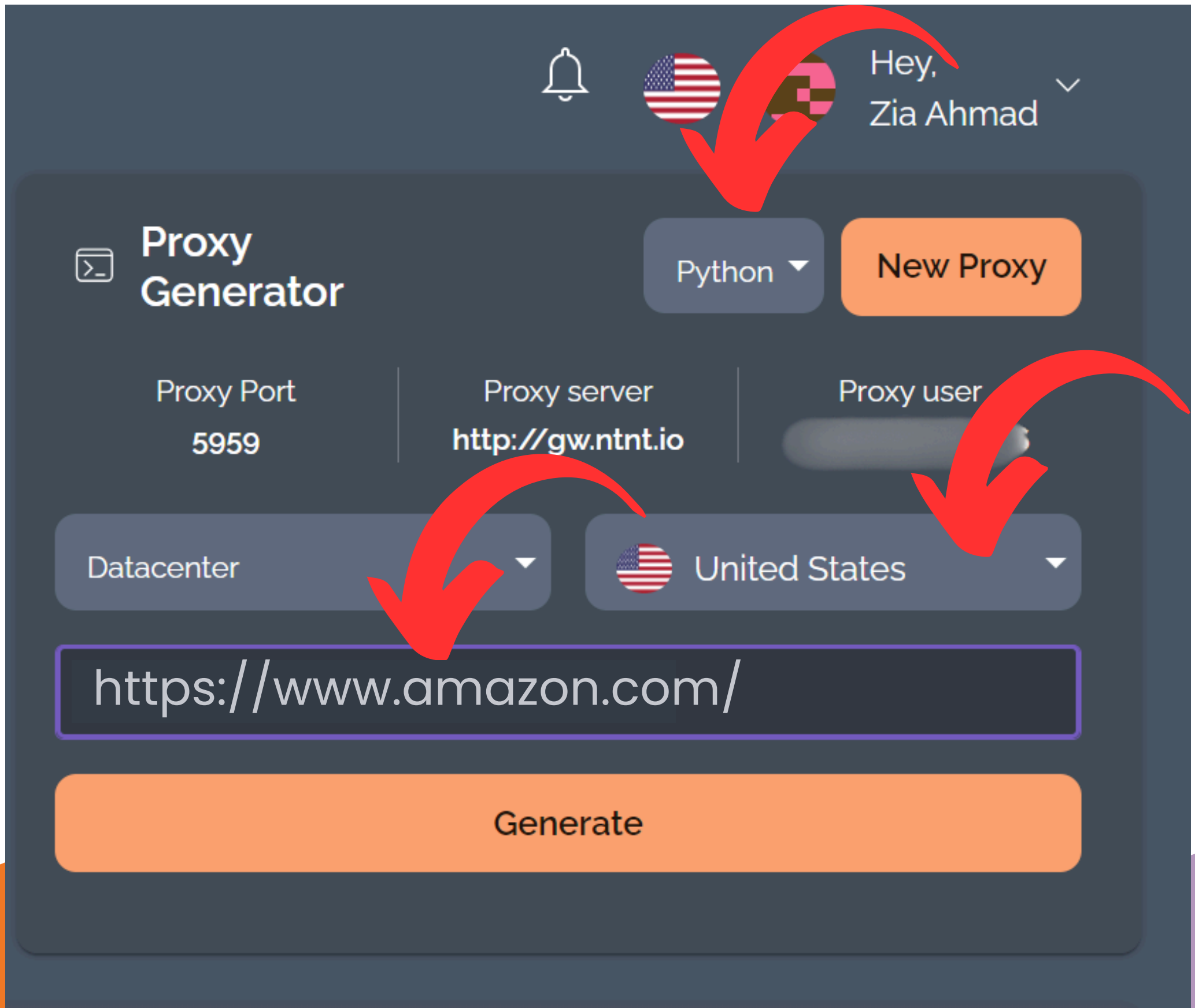
Sign up to:

https://portal.netnut.io/



Set up the Proxy

Pick your preferred Programming Language, Location & URL



- **Import Libraries**
- **Paste the Data that we got from NetNut.io**

```
from selenium import webdriver
from webdriver_manager.chrome import ChromeDriverManager
from bs4 import BeautifulSoup
import time
```

```
username = 'zi[illegible]S'
password = 'r8[illegible]OM'
server = 'gw.ntnt.io'
port = '5959'
proxy_url = f'http://{username}:{password}@{server}:{port}'
```

- **Setting up WebDrivers**
- **Hitting the URL of Amazon.com to get the data**

```
capabilities = webdriver.DesiredCapabilities.CHROME
options = webdriver.ChromeOptions()
options.add_argument(f'--proxy-server={proxy_url}')
driver = webdriver.Chrome(ChromeDriverManager().install(),
    chrome_options = options, desired_capabilities = capabilities)

keyword = "Cars"
url = f'https://www.amazon.com/'
driver.get(url)
time.sleep(3)
driver.find_element_by_xpath('//*[@id="b"]').click()
time.sleep(3)
driver.find_element_by_id("twotabsearchtextbox").send_keys(keyword)
driver.find_element_by_id('nav-search-submit-button').click()
```

- **Parsing the required Data**
- **Printing the**
  - **Product name**
  - **Price**

```
soup = BeautifulSoup(driver.page_source, 'html.parser')
post_cards = soup.find_all("div", "a-section a-spacing-small a-spacing-top-small")
data = []
for i in post_cards:
    dct = {}
    try:
        dct['Price'] = i.find(class_="a-offscreen").get_text()
        dct['Title'] = i.find(class_="a-size-medium a-color-base a-text-normal").get_text()
        data.append(dct)
    except:
        print("Didn't find the Text")

pd.DataFrame(data)
```

Sample Output

Price	Title
$3.79	Cars
$13.99	Hot Wheels Set of 10 Toy Cars & Trucks in 1:64...
$21.97	Mattel Disney and Pixar Cars Transforming Mack...
$23.00	Mattel Disney and Pixar Cars Mini Racers Set o...
$23.00	Matchbox Cars, 20-Pack of 1:64 Scale Die-Cast ...
$13.49	Mattel Disney and Pixar Cars Mini Racers 3-Pac...
$22.15	Mattel Disney and Pixar Cars Set of 5 Collecti...
$38.87	Carrera First Disney/Pixar Cars - Slot Car Rac...
$40.40	Mattel Disney and Pixar Cars Set of 10 Die-Cas...
$3.79	Cars 3 (Theatrical)
$3.79	Cars 2

Scraping Amazon.com using Python

Hit Follow!

Save for later!

netnut.io