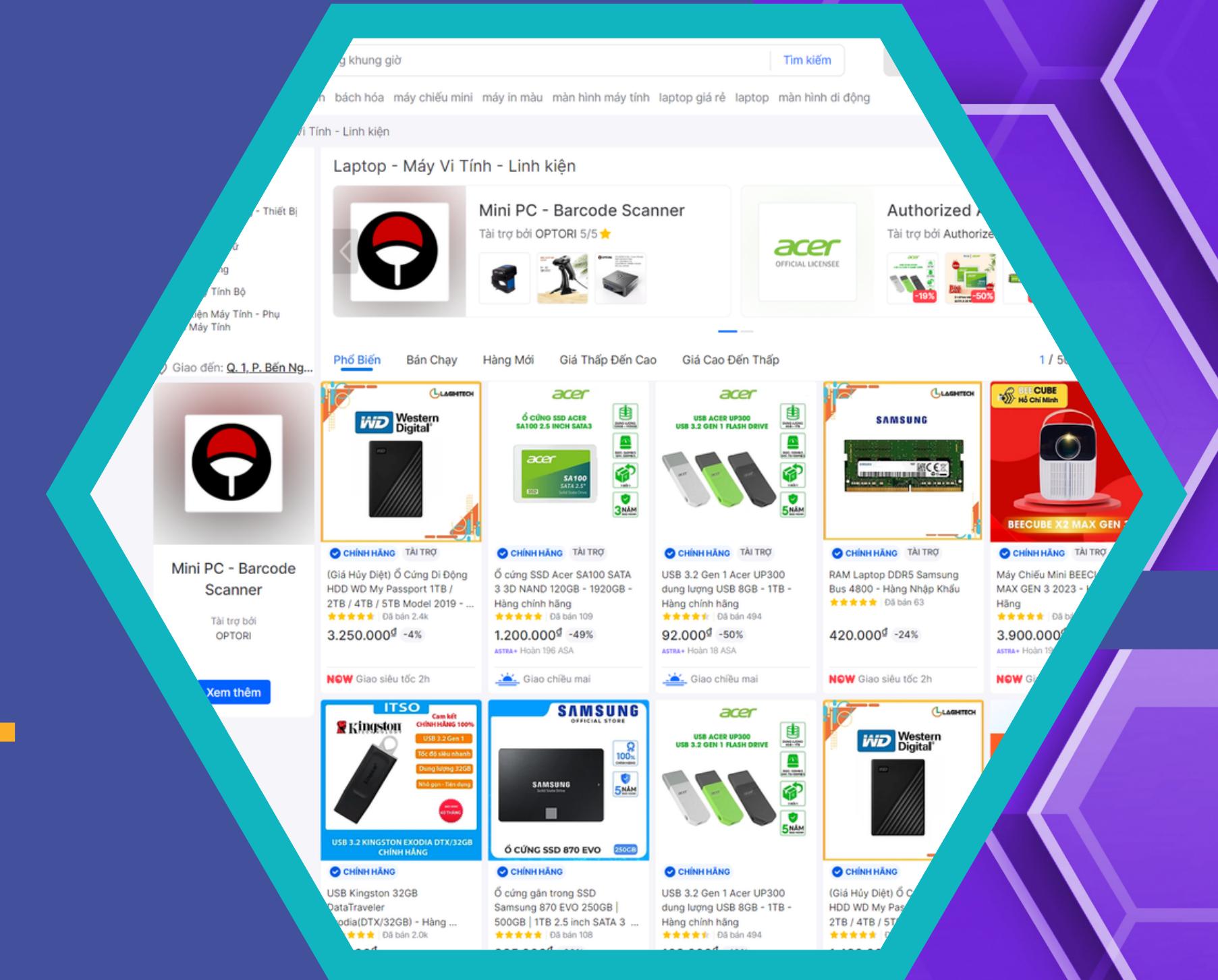
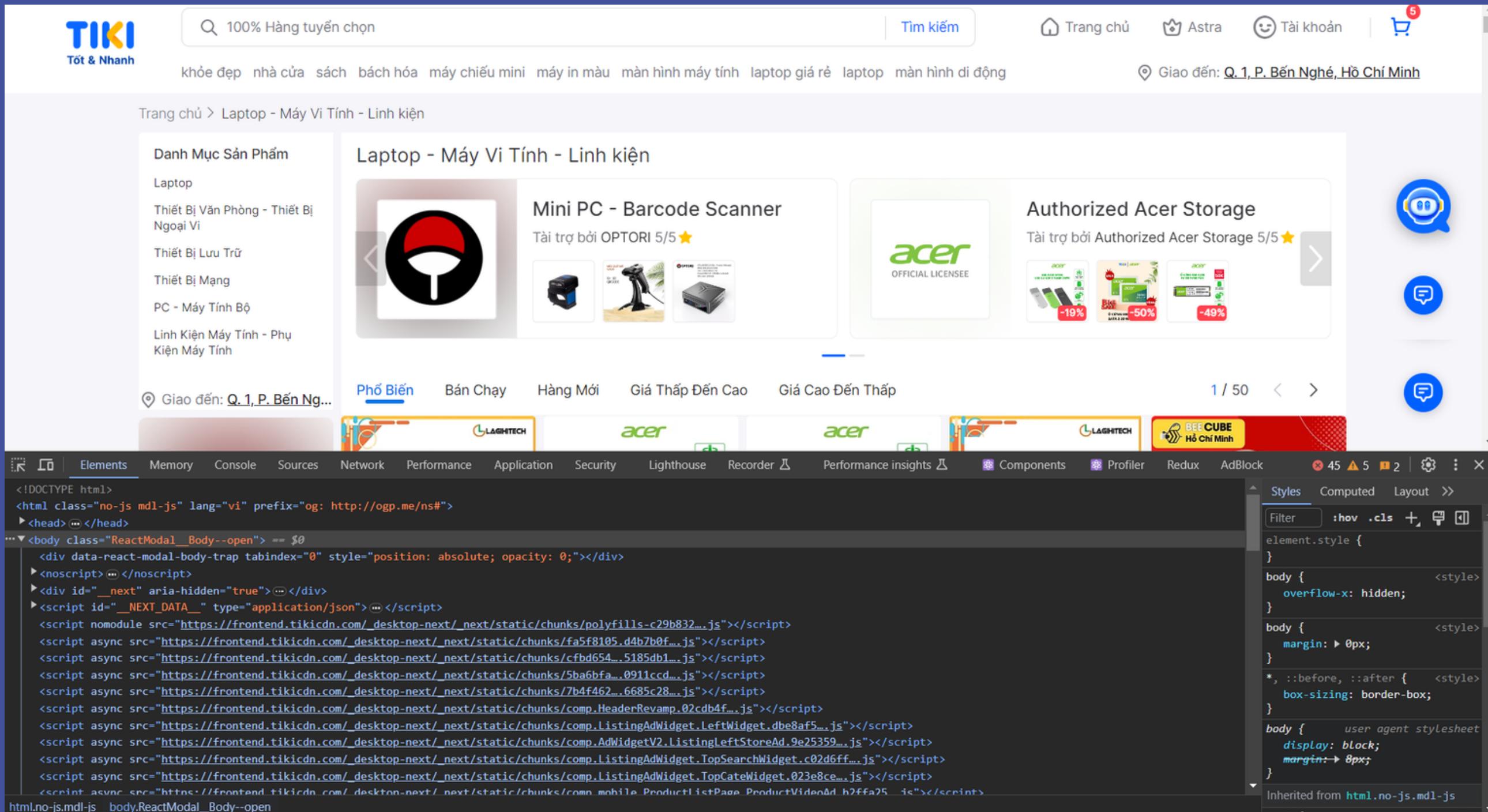
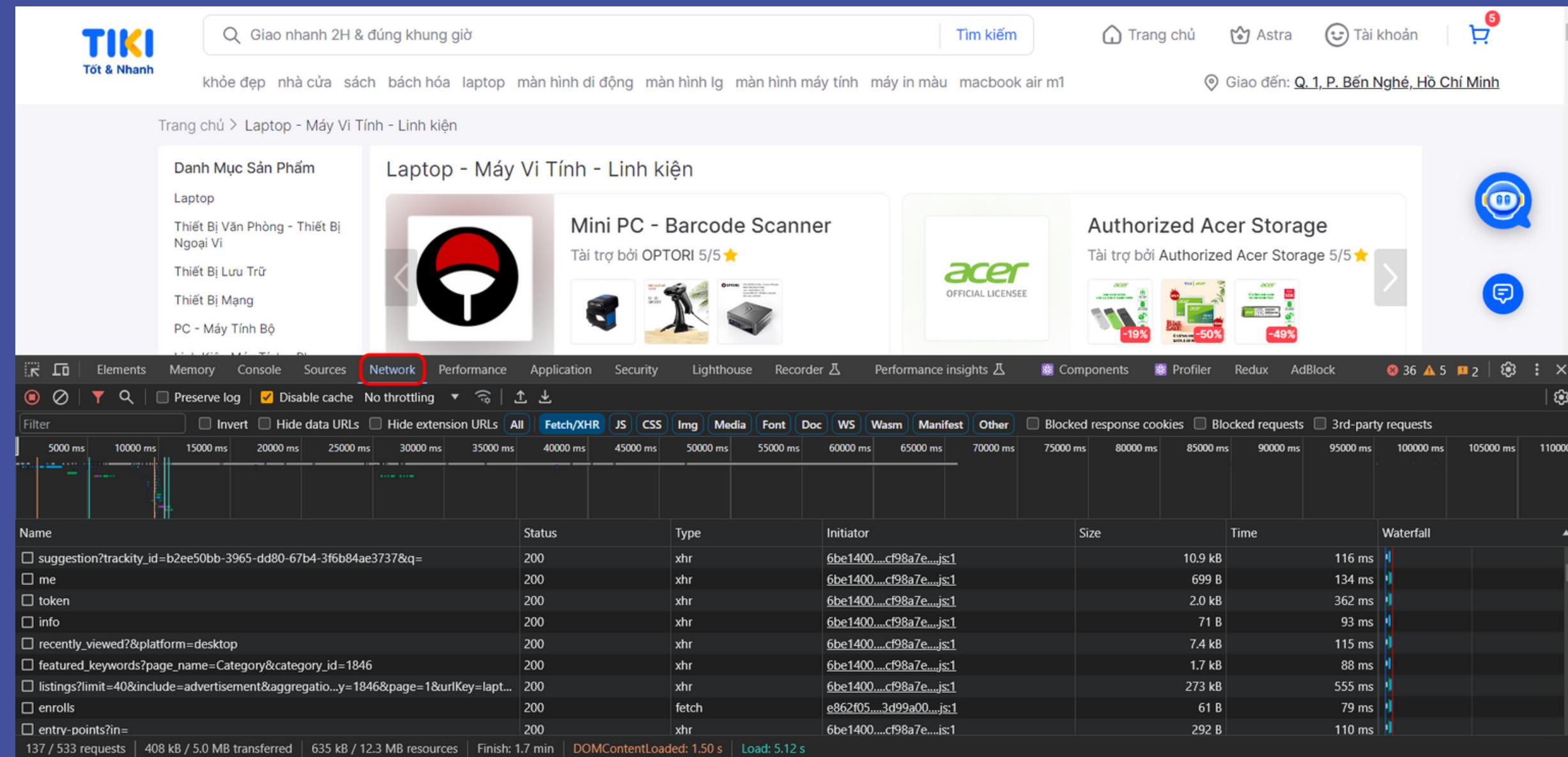


# TechX CRAWL DATA TIKI PRODUCT





CÁC BẠN CHỈ CẦN INSPECT Ở BẤT KÌ VỊ TRÍ  
NÀO TRÊN TRANG WEB LÀ ĐƯỢC.



TỪ TAB ELEMENT, CÁC BẠN DI CHUYỂN  
SANG TAB NETWORK TRÊN DEVTOOLS



The screenshot shows a browser developer tools window with the Network tab selected. The main area displays a shopping website with various products like SSDs and RAM. One product listing for a Western Digital My Passport drive has a red box around its description. Below the main content, the Network tab lists several API requests. One request, highlighted with a red box, is for a product listing with the URL `listings?limit=40&include=a..`. The response payload shows a JSON object with a "block" key containing a "code" of "tiki\_listing" and a "title" of "Tiki Listing". The "data" array contains a single item with an ID of 38323087 and a SKU of 9802621862644. The name of the product is listed as "(Giá Hủy Diệt) Ổ Cứng Di Động HDD WD My Passport 1TB / 2TB / 4TB / 5TB Model 2019 - Hàng Nhập Khẩu". The bottom of the developer tools shows the total number of requests (137/560) and the total size (408 kB / 5.1 MB).

ĐÂY LÀ API MÀ TIKI SỬ DỤNG ĐỂ LẤY DỮ  
LIỆU SẢN PHẨM TỪ SERVER



```
1 https://tiki.vn/api/personalish/v1/blocks/listings?limit=40&include=advertisement&aggregations=2&version=home-personalized&trackity_id=b2ee50bb-3965-dd80-67b4-3f6b84ae3737&category=1846&page=1&urlKey=laptop-may-vi-tinh-linh-kien
```



```
1 https://tiki.vn/api/personalish/v1/blocks/listings
```



```
1 limit=40&include=advertisement&aggregations=2&version=home-personalized&trackity_id=b2ee50bb-3965-dd80-67b4-3f6b84ae3737&category=1846&page=1&urlKey=laptop-may-vi-tinh-linh-kien
```



**API MÀ TIKI LẤY DỮ LIỆU  
SẢN PHẨM TỪ SERVER**



**ENDPOINT API CỦA TIKI**



**PARAM TRUYỀN  
VÀO ENDPOINT**



CRAWL DATA WITH PYTHON

CẤU HÌNH COOKIES,  
HEADER, PARAMS

params = {} ←

▼Query String Parameters [view source](#) [view URL-encoded](#)

- limit: 40
- include: advertisement
- aggregations: 2
- version: home-personalized
- trackity\_id: b2ee50bb-3965-dd80-67b4-3f6b84ae3737
- category: 1846
- page: 1
- urlKey: laptop-may-vi-tinh-linh-kien

## CÀI ĐẶT THƯ VIỆN REQUESTS ĐỂ GỬI REQUESTS API ĐẾN SERVER TIKI



1 pip install requests

Cú pháp gọi API: **import requests**

```
response = requests.get( 'url' ,headers, params)  
response.json().get('data')
```

## Lưu danh sách chứa data crawl về thành công

```
products = [ ]
```

```
for i in range(1, 50):
```

```
with open("products_tiki.json", "w", encoding="utf-8") as f:  
    json.dump(products, f, ensure_ascii=False, indent=4)
```



**CHÚNG TA SẼ LƯU TẮT CẢ CÁC SẢN PHẨM  
ĐÃ CRAWL VÀO FILE ĐỊNH DẠNG JSON :**



time.sleep(random.randrange(1, 3))



THIẾT LẬP THỜI GIAN CHỜ  
SAU MỖI LẦN GỬI REQUEST



TRÁNH TÌNH TRẠNG TIKI HIỂU NHẦM CHÚNG TA ĐANG DDOS  
ĐẾN TRANG TIKI, CHÚNG TA SẼ THIẾT LẬP THỜI GIAN CHỜ  
NGẪU NHIÊN GIỮA CÁC LẦN GỬI REQUEST.



# THÀNH QUẢ ĐẠT ĐƯỢC

```
You, 5 days ago • create file crawlDataTiki.py for crawl data tiki
1 [ 2 {
  3   "id": 170497385,
  4   "name": "Điện thoại Samsung Galaxy A53 5G (8GB/128GB)",
  5   "brand_name": "Samsung",
  6   "original_price": 9990000,
  7   "price": 6990000,
  8   "discount": 30,
  9   "thumbnail_url": "https://salt.tikicdn.com/cache/280x280/ts/product/77/10/74/
ec9638a316bf7acc3c91afe2c6af44e7.jpg",
10   "quantity_sold": {
11     "text": "Đã bán 621",
12     "value": 621
13   },
14 },
15 > ...
16 > ...
17 > ...
18 > ...
19 > ...
20 > ...
21 > ...
22 > ...
23 > ...
24 > ...
25 > ...
26 > ...
27 > ...
28 > ...
29 > ...
30 > ...
31 > ...
32 > ...
33 > ...
34 > ...
35 > ...
36 > ...
37 > ...
38 > ...
39 > ...
40 > ...
41 > ...
42 > ...
43 > ...
44 > ...
45 > ...
46 > ...
47 > ...
48 > ...
49 > ...
50 > ...
51 > ...
52 > ...
53 > ...
54 > ...
55 > ...
56 > ...
57 > ...
58 > ...
59 > ...
60 > ...
61 > ...
62 > ...
63 > ...
64 > ...
65 > ...
66 > ...
67 > ...
68 > ...
69 > ...
70 > ... },
```



# NHỮNG KHÓ KHĂN KHI CRAWL DATA



## Phân tích dự án

- **Cấm truy cập và giới hạn IP**
- **Cấm truy cập bởi robots.txt**
- **Hạn chế về tốc độ**
- **Thay đổi cấu trúc trang web**
- **Một số cấu hình nâng cao để gửi nhiều request trong một thời điểm**
- **Độ trễ và xử lý lỗi**



MỤC ĐÍCH CRAWL DATA TRONG  
BUỔI HỌC NÀY CHỈ PHỤC VỤ CHO  
VIỆC HỌC VÀ NGHIÊN CỨU VỀ CẤU  
TRÚC DỮ LIỆU VÀ GIẢI THUẬT.