

Coding Challenge: Build a Scalable and Undetectable API for Scraping TikTok Shop Product Data

Objective

Your task is to build a scalable and undetectable API that scrapes product data from **TikTok Shop**. The scraper must retrieve JSON response data from internal APIs used by TikTok Shop when:

- Searching products **by shop**.
- Searching products **by category**.
- Viewing a **product detail page**.

Target API Behaviours

Your scraper should be able to detect and extract JSON response payloads from:

- Product list API when searching by **shop name or ID**.
- Product list API when browsing by **category**.
- **Product details API** for individual products.

These endpoints are typically visible in the browser's DevTools > Network tab when interacting with TikTok Shop web pages.

Requirements:

1. Scraping Logic

Extract the **raw JSON** response from each internal API:

- Product list by shop.
- Product list by category.
- Product detail

Output should match the full JSON structure from TikTok's API responses.

Implement **anti-detection strategies**:

- Rotate **user agents**, headers, and Ips.
- Use **proxies** or **residential proxy pools**.
- Use **headless browsers** (e.g., Playwright, Puppeteer) if necessary.

2. API Development

Build a REST API that supports the following endpoints:

- GET /tiktok/search-by-shop?shopId={shop_id}
- GET /tiktok/search-by-category?categoryId={cat_id}
- GET /tiktok/product-detail?productId={product_id}

Each should return the raw API JSON response from TikTok Shop.

3. Tech Stack

- Use **JavaScript** to implement the scraper and API.
- **TypeScript** is a strong plus and will be considered a bonus.

4. Hosting

- Deploy your API and expose it publicly using **Ngrok** (or similar).
- Share the public URL so we can test it.

✅ Test Success Criteria

To pass this challenge, your API must:

- ✅ Successfully retrieve data from **at least 1000 TikTok Shop products**.
- ✅ Maintain **average latency ≤ 30 seconds per request**.
- ✅ Handle scraping under **light load with stability** for 1 hour.
- ✅ Maintain **≤ 5% error rate**.

Scraping & Proxy Notes

 **You may not use our proxies** — due to security concerns.

 You **must manage proxies yourself**.

 You are allowed to use **free trial proxy providers**.

Deliverables

Hosted API URL (via Ngrok or similar).

GitHub repo with:

Source code.

README.md with:

Setup instructions

Example API usage

Proxy configuration details.

Scraper techniques (e.g., bypass strategies).