

# Tiki-Data-Crawling Tool Instruction

Dacekey - Le Nguyen

February 17, 2026

## 1 Introduction

In the context of data collection for analytical and predictive purposes, an important question arises: how can one obtain a high-quality raw dataset—one that is sufficiently well-structured and reliable—to enable effective processing in subsequent stages of the workflow?

For this project, I developed a custom tool to scrape data from an e-commerce platform, Tiki. In this article, I will provide a step-by-step guide on how to use this tool.

## 2 Where Can the Data Be Found?

**Step 1:** Navigate to the [Tiki](#) website.

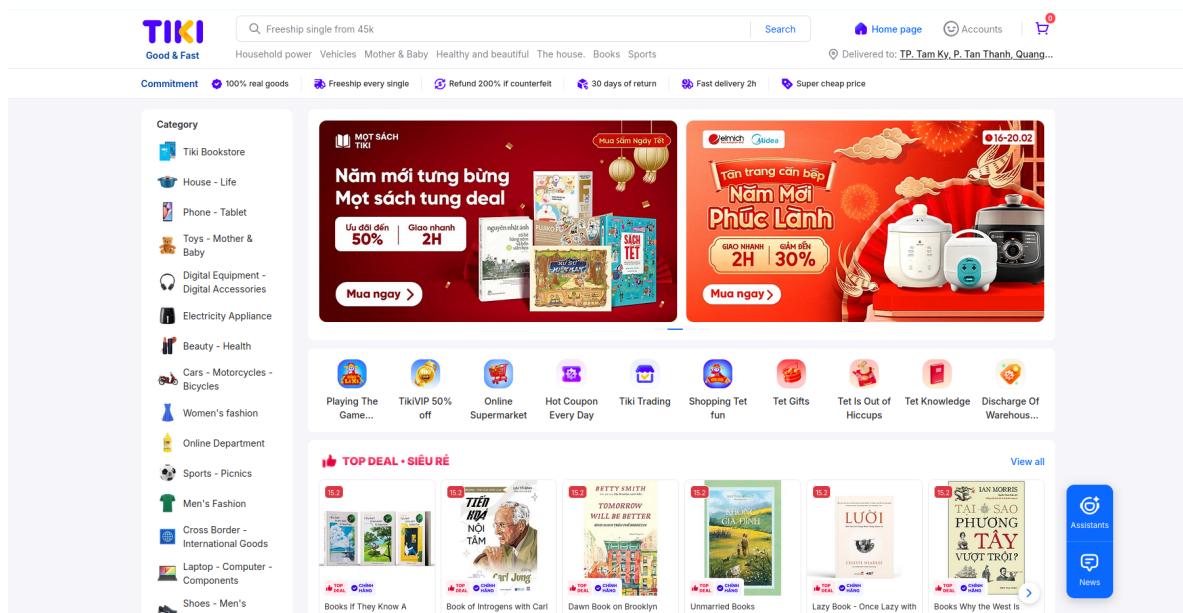


Figure 1: Tiki Homepage

**Step 2:** Select the target category for data collection.

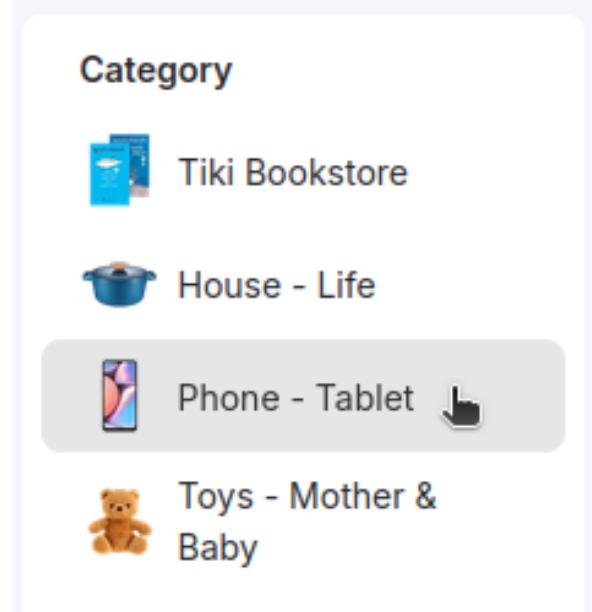


Figure 2: Tiki Category

**Step 3:** Access the browser's developer tools by pressing **F12**, and subsequently reload the webpage using **Ctrl + R**.

The screenshot shows the Network tab of the browser's developer tools. It lists several requests made to the domain `tiki.vn`. The requests include various Google Tag Manager scripts, analytics.google.com, and trackby.tiki.vn. The table includes columns for Status, Method, Domain, File, Initiator, Type, Transferred, and Size. The total number of requests is 75, with a total transferred size of 394.93 kB.

Status	Method	Domain	File	Initiator	Type	Transferred	Size
200	POST	api.amplitude.com	/	xhr	html	214 B	0 ms
200	OPTIONS	api.amplitude.com	/	xhr	plain	300 B	0 B   201 ms
200	POST	www.google.com	collectfmr=0&en=page_viewId=https://tiki.vn/dien-thoai-may-tinh-bang/c17898/scrsrc=www.googletagmanager.co...	gtag.js[243] (fetch)	plain	536 B	0 B   220 ms
200	POST	www.google.com	collectfmr=0&en=page_viewId=https://tiki.vn/dien-thoai-may-tinh-bang/c17898/scrsrc=www.googletagmanager.co...	gtag.js[116] (fetch)	plain	536 B	0 B   99 ms
200	POST	www.google.com	collectfmr=0&en=page_viewId=https://tiki.vn/dien-thoai-may-tinh-bang/c17898/scrsrc=www.googletagmanager.co...	gtag.js[171] (fetch)	plain	536 B	0 B   99 ms
200	POST	www.google.com	collectfmr=0&en=page_viewId=https://tiki.vn/dien-thoai-may-tinh-bang/c17898/scrsrc=www.googletagmanager.co...	gtag.js[171] (fetch)	plain	536 B	0 B   99 ms
200	POST	www.google.com	collectfmr=0&en=page_viewId=https://tiki.vn/dien-thoai-may-tinh-bang/c17898/scrsrc=www.googletagmanager.co...	gtag.js[190] (fetch)	plain	536 B	0 B   99 ms
200	POST	www.google.com	collectfmr=0&en=page_viewId=https://tiki.vn/dien-thoai-may-tinh-bang/c17898/scrsrc=www.googletagmanager.co...	gtag.js[190] (fetch)	plain	536 B	0 B   99 ms
200	GET	ep1.adtradicallquality.google	soder?sv=200&ot=gata&v=20260711&id=mv0j+4289110136468293	xhr	json	14.04 kB	17.90 kB   101 ms
200	POST	analytics.google.com	collectv=204&id=G_S9GAR1RQ2&gtm=45P6z6l9v16717717za200z91471771716_p=177132824941&gt=1313[259] (fetch)	gtag.js[259] (fetch)	plain	811 B	0 B   44 ms
200	GET	tiki.vn	me	xhr	json	843 B	745 B   1 ms
200	POST	trackby.tiki.vn	single	xhr	json	982 B	22 B   159 ms
200	OPTIONS	trackby.tiki.vn	single	xhr	plain	943 B	0 B   1 ms
200	GET	tiki.vn	GET /tiki/vn/dien-thoai-may-tinh-bang/c17898/scrsrc=www.googletagmanager.com?id=5209725	xhr	json	843 B	745 B   155 ms
200	POST	trackby.tiki.vn	single	xhr	json	982 B	22 B   159 ms
200	OPTIONS	trackby.tiki.vn	single	xhr	plain	943 B	0 B   1 ms

15 requests 394.93 kB / 438.18 kB transferred | Finish: 38.33 s DOMContentLoaded: 849 ms load: 3.13 s

Figure 3: Browser's Developer Tools

**Step 4:** Identify the file that contains the data corresponding to the selected category.

Figure 4: Json File

**Note:** Ensure that you are in the **Network** tab. Navigate to **XHR**, then select **Response** to verify whether any relevant data is returned. If necessary, enable **Disable Cache**.

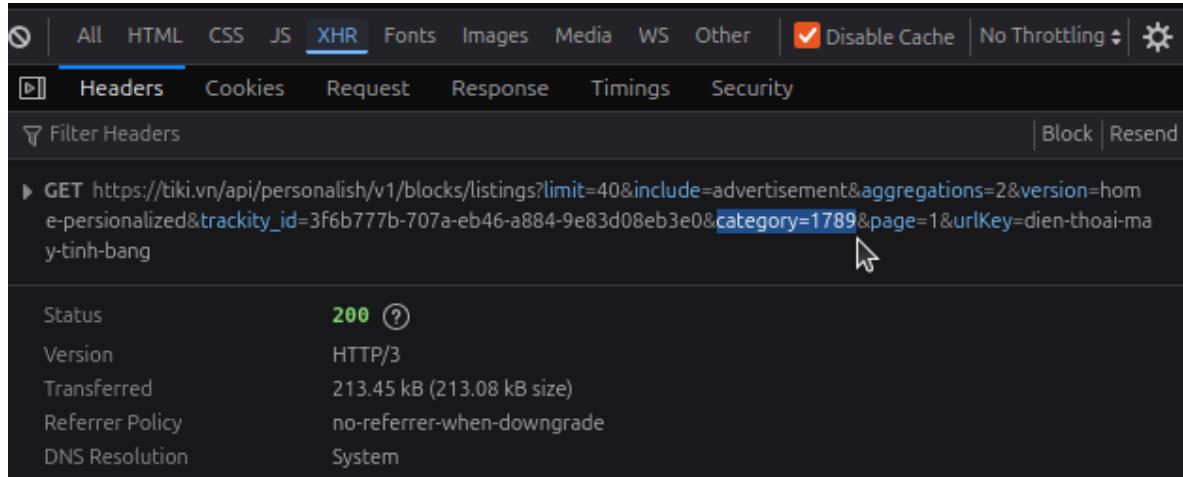
**Step 5:** Navigate to **Headers** and inspect the **URL**.

Block | Resend

▶ GET https://tiki.vn/api/personalish/v1/blocks/listings?limit=40&include=advertisement&aggregations=2&version=home-personalized&trackity_id=3f6b777b-707a-eb46-a884-9e83d08eb3e0&category=1789&page=1&urlKey=dien-thoai-may-tinh-bang	Block   Resend
Status	200 ⓘ
Version	HTTP/3
Transferred	212.01 kB (211.64 kB size)
Referrer Policy	no-referrer-when-downgrade
DNS Resolution	System

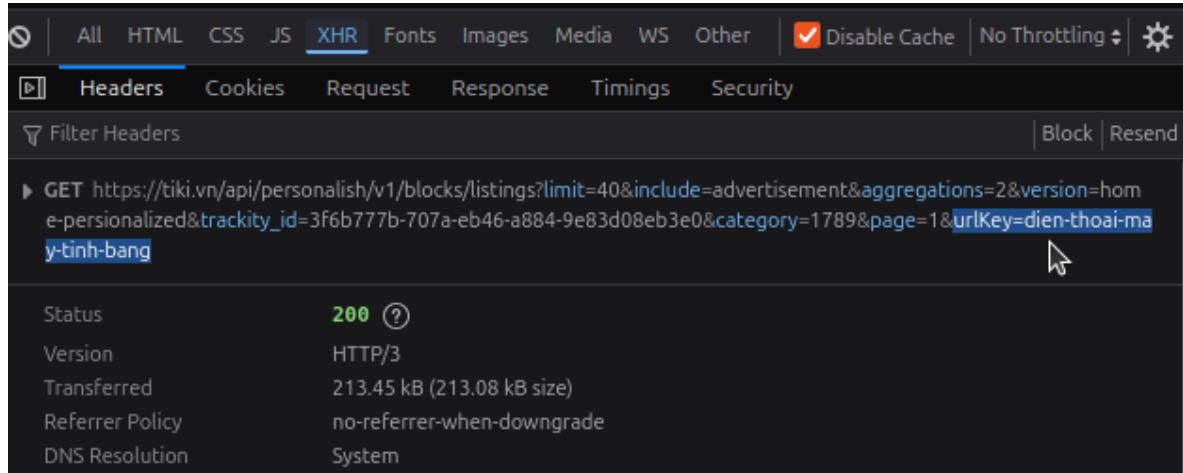
Figure 5: Header's URL

**Step 6:** Extract the corresponding values of **category** and **urlkey**.



The screenshot shows the NetworkMiner tool interface. The 'XHR' tab is selected. A network request is listed: GET https://tiki.vn/api/personalish/v1/blocks/listings?limit=40&include=advertisement&aggregations=2&version=hom e-personalized&trackity\_id=3f6b777b-707a-eb46-a884-9e83d08eb3e0&**category=1789**&page=1&urlKey=dien-thoai-ma y-tinh-bang. The 'category' value is highlighted in blue. Below the request, the response headers are displayed, including Status (200), Version (HTTP/3), Transferred (213.45 kB), Referrer Policy (no-referrer-when-downgrade), and DNS Resolution (System).

Figure 6: Category's Number



The screenshot shows the NetworkMiner tool interface. The 'XHR' tab is selected. A network request is listed: GET https://tiki.vn/api/personalish/v1/blocks/listings?limit=40&include=advertisement&aggregations=2&version=hom e-personalized&trackity\_id=3f6b777b-707a-eb46-a884-9e83d08eb3e0&**category=1789**&page=1&urlKey=dien-thoai-ma y-tinh-bang. The 'urlKey' value is highlighted in blue. Below the request, the response headers are displayed, including Status (200), Version (HTTP/3), Transferred (213.45 kB), Referrer Policy (no-referrer-when-downgrade), and DNS Resolution (System).

Figure 7: Category's Name

**Note:** In this example, the values '1798' and 'dien-thoai-may-tinh-bang' will be used for data collection.

**Step 7:** Go to file B\_TikiFinalCrawlingTool.py.

```
# =====
# MODIFY HERE
# 1. Choose Tiki category
category_name = "nha-sach-tiki"
category_numb = "8322"
filename_output = "tiki_library"
# 2. Choose number of pages
START_PAGE = 1
END_PAGE = 2
# =====
```

Figure 8: Python File

**Note:**

- Modify the `category_name` to ‘dien-thoai-may-tinh-bang’ and set the `category_num` to ‘1798’.
- Specify the desired name of the exported dataset by updating the `filename_output` parameter.
- The `START_PAGE` parameter is set to 1 by default.
- The total number of pages to be collected can be configured by adjusting the value of `END_PAGE`.

**Step 8:** Execute the script and verify the generated dataset.