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?
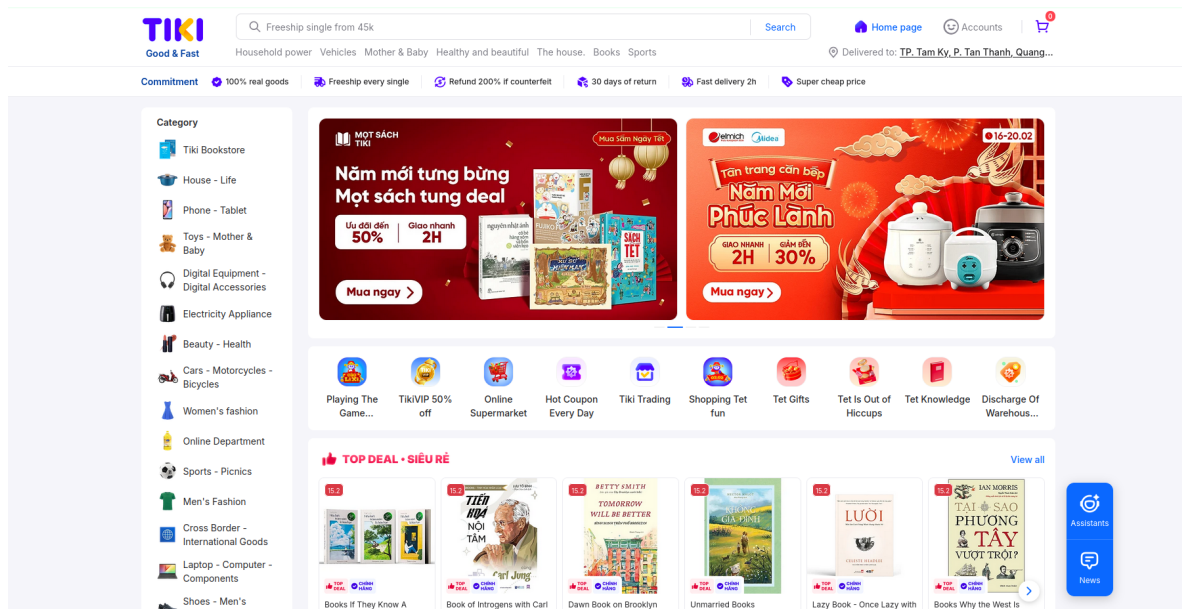
Step 1: Navigate to the Tiki website.

Figure 1: Tiki Homepage

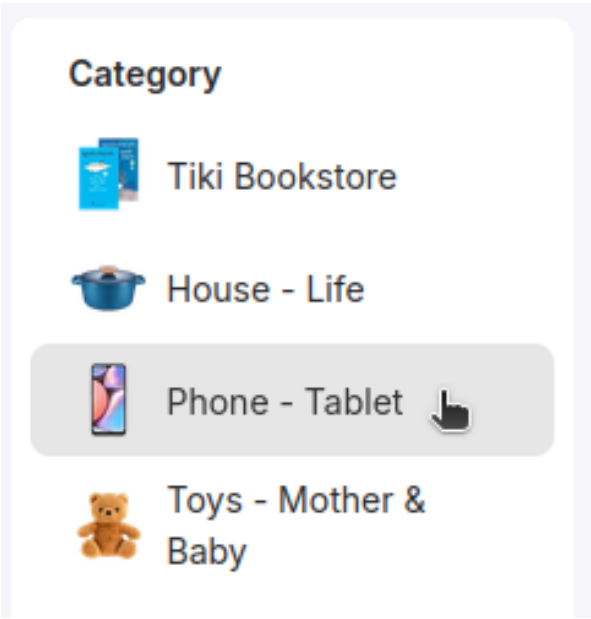**Step 2**: Select the target category for data collection.

Figure 2: Tiki Category

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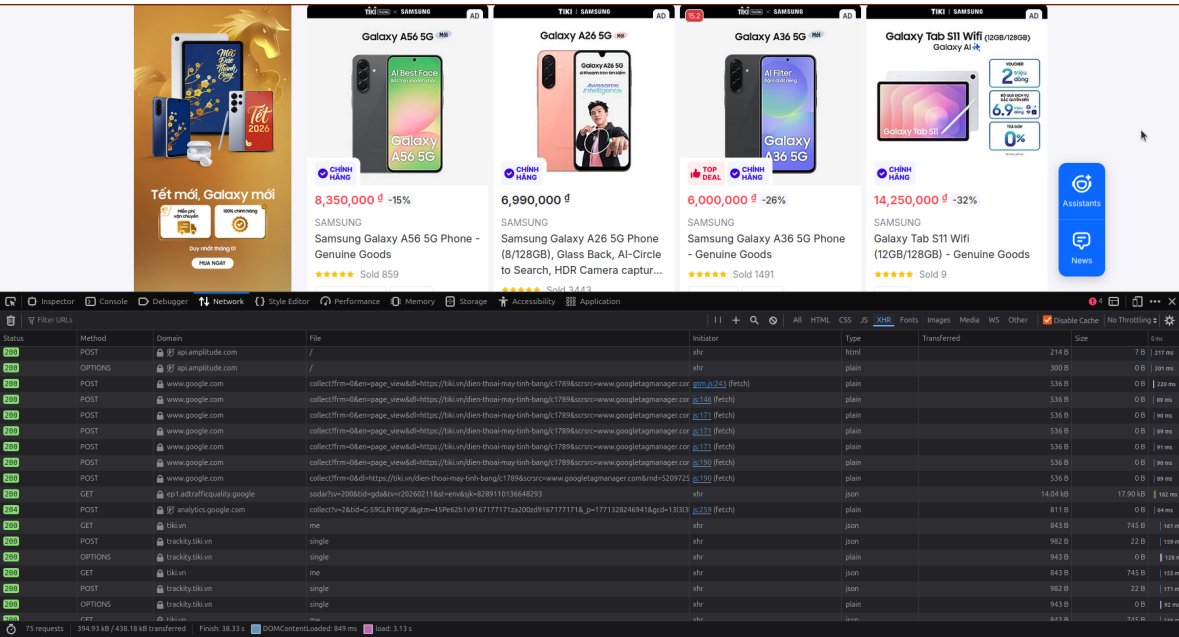

Figure 3: Browser's Developer Tools

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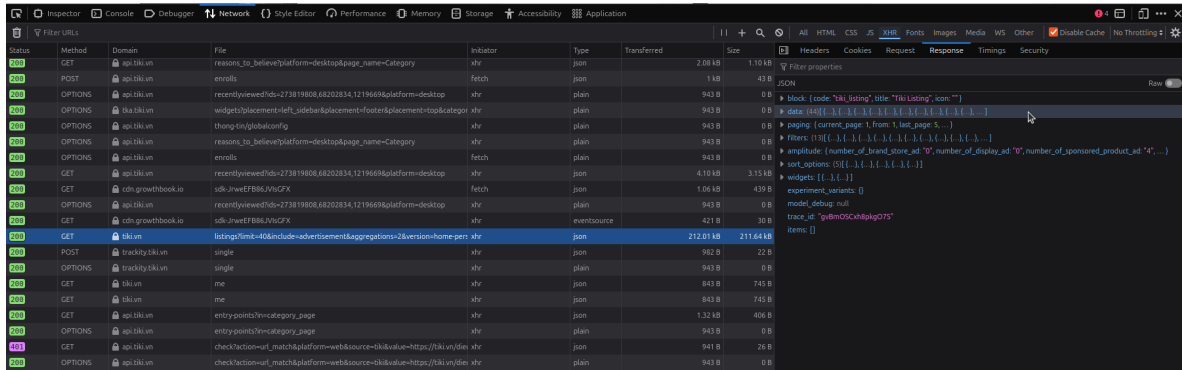

Figure 4: Json File

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**.

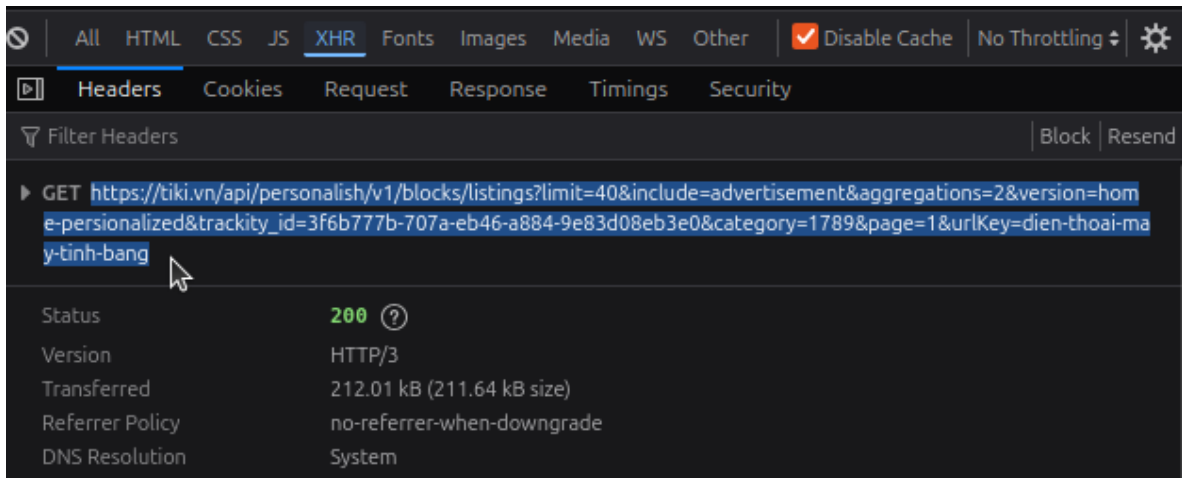**Step 5**: Navigate to **Headers** and inspect the **URL**.

Figure 5: Header's URL

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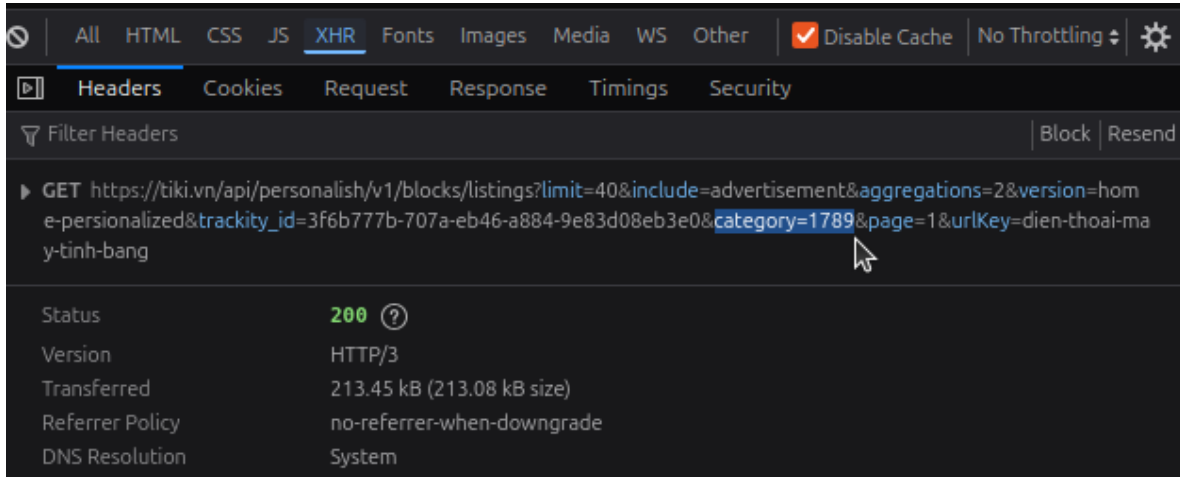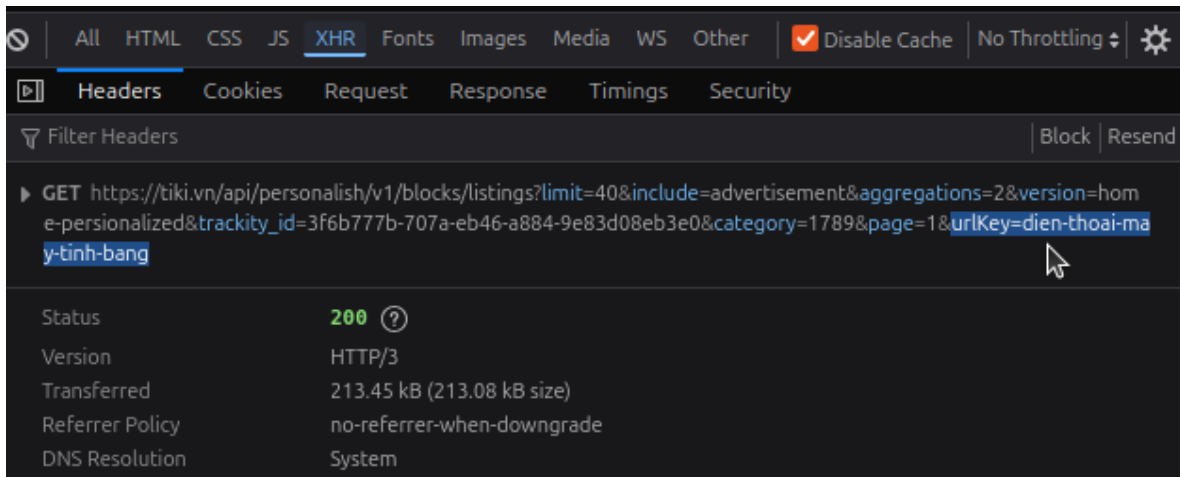

Figure 6: Category's Number

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.