

The background of the slide features a city skyline. On the left, a tall, slender skyscraper is under construction, its top section still skeletal. To its right, a cluster of modern high-rise buildings with glass facades stands against a sky transitioning from a soft pink to a pale blue. The overall aesthetic is modern and urban.

DATA SCIENCE

PROGRESS REPORT MIDTERM

MEMBERS:

- 1. HOANG PHUOC GIA NGUYEN - 20127574**
- 2. NGUYEN THIEN NHAN - 20127265**
- 3. NGUYEN MANH CUONG - 20127456**



CONTENT

01

Overview

02

Crawl Data

03

Preprocessing


04

Build Model

05

Evaluation

OVERVIEW




Tốt & Nhanh


[Tìm kiếm](#)


[Trang chủ](#)


[Tài khoản](#)


Giao đến: [Q. 10, P. 12, Hồ Chí Minh](#)


 [Ô tô - Xe Máy - Xe Đạp](#)


 [Thời trang nữ](#)


 [Bách Hóa Online](#)


 [Thể Thao - Thể Ngoại](#)


 [Thời trang nam](#)


 [Cross Border - Hàng Quốc Tế](#)


 [Laptop - Máy Vi Tính - Linh kiện](#)


 [Giày - Dép nam](#)


 [Điện Tử - Điện Lạnh](#)

 [Giày - Dép nữ](#)


 [Máy Ảnh - Máy Quay Phim](#)


 [Phụ kiện thời trang](#)

 [HOT NGON](#)





Công nghệ Tiki chọn Giảm tới 49%








Unilever chăm sóc toàn diện Coupon 8%





[Tiki Best](#)


[Nhập khẩu chính hãng](#)


[Khuyến mãi](#)


[Sản phẩm mới](#)

[Nhà Sách Tiki](#)

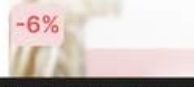
[Nhà Cửa - Đời Sống](#)


[Thiết Bị Số - Phụ Kiện Số](#)


[Trợ lý](#)


[Tin mới](#)


Giá tốt hôm nay 00 : 43 : 07














[Xem tất](#)

https://tiki.vn/khuyen-mai/tech-arena?itm_campaign=HMP_YPD_TKA_BNA_UNK_AL...

CRAWL DATA

The image shows a web browser displaying a clothing store page with various items like shorts and jackets. Below the page, the browser's developer tools are open, specifically the Network tab. A red box highlights the 'Fetch/XHR' filter, and another red box highlights a specific request in the list: 'listings?limit=40&include=advertisement&aggregatio...'. The details for this request are shown on the right, including the Request URL, Request Method (GET), Status Code (200 OK), Remote Address, and Referrer Policy.

Đồ bơi - Đồ đi biển nam
Quần short nam

Giao đến: Q.10, P.12, H...

Dịch vụ
☐ NOW Giao Siêu Tốc 2H

Đánh giá
★★★★★ từ 5 sao
★★★★☆ từ 4 sao

Set 2 quần đùi nam, quần Short
Combo 2 quần đùi nam, quần
Combo 2 Quần shorts thể thao
Áo Khoác Dù Nam chống nước
Áo chống nắng nam cao cấp,

CHÍNH HÃNG TÀI TRỢ

Trợ lý
Tin mới

Network

Filter
☐ 3rd-party requests
☐ Big request rows
☒ Overview
☐ Invert
☐ Hide data URLs
☐ Hide extension URLs
All Fetch/XHR Doc CSS JS Font Img Media Manifest WS Wasm Other
☐ Blocked response cookies
☐ Blocked requests

50000 ms 100000 ms 150000 ms 200000 ms 250000 ms 300000 ms 350000 ms 400000 ms 450000 ms 500000 ms 550000 ms 600000 ms 650000 ms

Name
total
info
listings?limit=40&include=advertisement&aggregatio...
sdk-JrweEF800vVix3FX
tracking
single
single
api.amolitude.com

190 / 711 requests 2.3 MB / 2.8 MB transferred 2.4 MB / 16.7

Headers Payload Preview Response Initiator Timing Cookies

General
Request URL: https://tiki.vn/api/personalish/v1/blocks/listings?limit=40&include=advertisement&aggregations=2&version=home-personalized&trackity_id=f78e9252-5f5d-c4cb-9bbd-1d288df6d76d&category=915&page=1&urlKey=thoi-trang-nam
Request Method: GET
Status Code: 200 OK
Remote Address: 35.186.195.157:443
Referrer Policy: no-referrer-when-downgrade

Response Headers

CRAWL DATA

```
1 {
2   "block": {
3     "code": "tiki_listing",
4     "title": "Tiki Listing",
5     "icon": ""
6   },
7   "data": [
8     {
9       "id": 99968881,
10      "sku": "3392225491485",
11      "name": "Set 2 quần đùi nam, quần Short Giò nam thể thao Basic trẻ trung năng động, thoáng mát co giãn 4 chiều MRM Manlywear",
12      "url_key": "set-2-qua-n-du-i-nam-qua-n-short-gio-nam-the-thao-basic-tre-trung-nang-do-ng-thoa-ng-ma-t-co-gia-n-4-chie-u-mrm-manlywear-p99968881",
13      "url_path": "set-2-qua-n-du-i-nam-qua-n-short-gio-nam-the-thao-basic-tre-trung-nang-do-ng-thoa-ng-ma-t-co-gia-n-4-chie-u-mrm-manlywear-p99968881.html?spid=99968992",
14      "type": "",
15      "author_name": "",
16      "book_cover": null,
17      "brand_name": "MRM Manlywear",
18      "short_description": "",
19      "price": 229500,
20      "list_price": 0,
21      "badges": [],
22      "badges_new": [
23
```

Cách crawl:

- Crawl id tất cả sản phẩm của một category và lưu vào product_id_ncds.csv
- Dùng danh sách id đó để crawl thông tin của từng sản phẩm trong một category

CRAWL PRODUCT ID

```
> cookies = { ...

headers = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/122.0.0.0 Safari/537.36',
    'Accept': 'application/json, text/plain, */*',
    'Accept-Language': 'vi-VN,vi;q=0.8,en-US;q=0.5,en;q=0.3',
    'Referer': 'https://tiki.vn/?src=header_tiki',
    'x-guest-token': '8jWSuID8b2NGVzr6hsUZxpkP1FRin71Y',
    'Connection': 'keep-alive',
    'TE': 'Trailers',
}

params = {
    'limit': '60',
    'include': 'sale-attrs,badges,product_links,brand,category,stock_item,advertisement',
    'aggregations': '1',
    'trackity_id': 'f78e9252-5f5d-c4cb-9bbd-1d288df6d76d',
    'category': '976',
    'page': '1',
    'src': 'c976',
    'urlKey': 'tui-thoi-trang-nu',
}
```

CRAWL PRODUCT ID

AutoSave Off product_id_ncd... Saved to this PC Search NGUYỄN THIÊN NHÂN

File Home Insert Page Layout Formulas Data Review View Automate Help Foxit PDF Comments Share

Clipboard Font Alignment Number Styles Cells Editing Add-ins Analyze Data

A1 : X ✓ fx Id

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Id																			
2	1.93E+08																			
3	1.87E+08																			
4	1.9E+08																			
5	2.08E+08																			
6	1.04E+08																			
7	1.94E+08																			
8	1.91E+08																			
9	1.6E+08																			
10	89428878																			
11	1.55E+08																			
12	97707630																			
13	2E+08																			
14	2.72E+08																			
15	1.99E+08																			
16	2.75E+08																			
17	1.92E+08																			
18	89773490																			
19	1.16E+08																			
20	1.68E+08																			
21	1.17E+08																			
22	1.86E+08																			
23	1.62E+08																			
24	1.89E+08																			
25	1.77E+08																			
26	1.99E+08																			
27	95289176																			
28	77937135																			
29	73384998																			
30	1.2E+08																			

product_id_ncds

Ready Accessibility: Unavailable 100%

CRAWL SAN PHAM

AutoSave Off | File Home Insert Page Layout Formulas Data Review View Automate Help Foxit PDF | Search | NGUYỄN THIÊN NHÂN | Comments | Share

Clipboard | Font | Alignment | Number | Styles | Cells | Editing | Add-ins

R10

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	id	sku	name	short_desc	price	list_price	discount	discount_r	review_count	rating_aver	inventory	quantity	scbrand_id	brand_name	spid	seller_id				
2	1.35E+08	6.18E+12	Giày da nam	Giày da nam	663000	970000	307000	32	41	4.7	available	91	4738431	Bá-Vi Leath	1.35E+08	181498				
3	1.35E+08	9.99E+12	Giày da nam	Giày da nam	563000	827000	264000	32	90	4.8	available	232	4738431	Bá-Vi Leath	1.35E+08	181498				
4	1.35E+08	4.15E+12	Giày da nam	Giày da nam	568000	985000	417000	42	165	4.7	available	438	4738431	Bá-Vi Leath	1.86E+08	181498				
5	1.92E+08	9.18E+12	Giày da nam	...	565000	840000	275000	33	15	4.6	available	62	4738431	Bá-Vi Leath	1.92E+08	181498				
6	2.74E+08	8.98E+12	Giày da nam	Giày da nam	420000	840000	420000	50	2	5	available	8	153494	Bee Gee	2.13E+08	975				
7	2.57E+08	4.54E+12	Keo dán	Keo dán	27000	35000	8000	23	965	4.6	available	5865	4995383	Ximo	2.57E+08	196837				
8	1.37E+08	4.05E+12	Mũ bảo vệ	Mũ bảo vệ	49500	99000	49500	50	90	4.6	available	838	49369	Pierre Card	1.37E+08	15446				
9	66619316	9.56E+12	Giày da nam	Giày da nam	359000	450000	91000	20	439	4.7	available	1075	4267275	ZAVAS	66619324	11871				
10	2.15E+08	6.2E+12	Chai xịt	Chai xịt	69000	69000	0	0	16	4.8	available	155	4995383	Ximo	2.53E+08	1028				
11	44969160	2.41E+12	Ốp lưng	Ốp lưng	36000	70000	34000	49	210	4.6	available	985	496813	buybox	44969162	72338				
12	2.74E+08	6.58E+12	Giày da nam	Giày da nam	349000	349000	0	0	1	5	available	2	4993639	THE BILY	2.74E+08	196712				
13	1.35E+08	2.1E+12	Giày da nam	Giày da nam	568000	650000	82000	13	38	5	available	90	4738431	Bá-Vi Leath	1.35E+08	181498				
14	2.48E+08	4.89E+12	Giày da nam	THẢNH TIN	598000	650000	52000	8	0	0	available	1	231125	Trần Anh	2.48E+08	11395				
15	1.92E+08	8.36E+12	Giày da nam	...	555000	840000	285000	34	19	4.5	available	59	4738431	Bá-Vi Leath	1.92E+08	181498				
16	7386965	1.92E+12	Balo	Balo	470000	470000	0	0	178	4.6	available	1531	243661	Crep Protec	14076477	9652				
17	13744574	1.66E+12	Combo 3	Combo 3	23000	26299	3299	13	75	4.3	available	379	111461	OEM	65180194	106273				
18	14885304	2.16E+12	Túi xách	Thần	16500	16500	0	0	93	4.3	available	434	111461	OEM	72309284	64160				
19	42920863	2.95E+12	Lốp xe	Ngân	50000	85000	35000	41	381	4.6	available	2133	496813	buybox	42920865	72338				
20	43035743	8.89E+12	Lốp xe	VACE	72500	145000	72500	50	198	4.6	available	1406	496813	buybox	43035751	72338				
21	45224922	7.94E+12	Lốp xe	VACE	42500	85000	42500	50	143	4.5	available	947	496813	buybox	45224930	72338				
22	2.48E+08	8.8E+12	Giày da nam	THẢNH TIN	506000	550000	44000	8	1	5	available	4	231125	Trần Anh	2.48E+08	11395				
23	1.92E+08	5.59E+12	Dép nam	...	299000	450000	151000	34	20	4.8	available	54	4738431	Bá-Vi Leath	1.92E+08	181498				
24	1.35E+08	5.28E+12	Giày da nam	Giày da nam	660000	990000	330000	33	30	4.6	available	80	4738431	Bá-Vi Leath	1.35E+08	181498				
25	1.35E+08	6.2E+12	Giày da nam	Giày da nam	568000	840000	272000	32	57	4.7	available	131	4738431	Bá-Vi Leath	1.35E+08	181498				
26	1.51E+08	4.53E+12	Giày da nam	Giày da nam	339000	339000	0	0	19	4.6	available	66	4993639	THE BILY	1.89E+08	196712				
27	66620209	7.95E+12	Giày da nam	Giày da nam	279000	450000	171000	38	67	4.7	available	179	4267275	ZAVAS	66620211	11871				
28	1.08E+08	1.56E+12	Balo	THẢNH TIN	39000	39000	0	0	2	5	available	33	4995383	Ximo	2.53E+08	1028				
29	49278142	9.71E+12	Xốp	Balo	86000	86000	0	0	129	4.8	available	561	496813	buybox	49278144	72338				
30	65778681	9.9E+12	Lốp xe	VACE	29000	45000	16000	36	132	4.5	available	683	496813	buybox	65778683	72338				

giay_dep_nam

Select destination and press ENTER or choose Paste | 100%

CRAWL SO SAO DANH GIA

AutoSave Off | File Home Insert Page Layout Formulas Data Review View Automate Help Foxit PDF | Search | NGUYỄN THIÊN NHÂN NT

Clipboard | Font | Alignment | Number | Styles | Cells | Editing | Add-ins | Analyze Data

Comments | Share

A1 | fx | product_id

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	product_id	seller_id	spid	1_star_cou	1_star_perc	2_star_cou	2_star_perc	3_star_cou	3_star_perc	4_star_cou	4_star_perc	5_star_cou	5_star_perc	rating_aver	reviews_count					
2	2.53E+08	11395	2.53E+08	0	0	0	0	0	0	1	50	1	50	4.5	2					
3	2.53E+08	11395	2.53E+08	0	0	0	0	0	0	0	0	1	100	5	1					
4	73120993	11395	73120995	0	0	0	0	0	0	0	0	0	0	0	0					
5	2.53E+08	11395	2.53E+08	0	0	0	0	0	0	0	0	0	0	0	0					
6	2.73E+08	975	2.73E+08	0	0	0	0	0	0	0	0	2	100	5	2					
7	43118520	72338	43118522	3	1	1	1	5	3	37	22	123	73	4.6	169					
8	1.09E+08	179250	1.09E+08	5	2	1	1	6	3	45	22	143	72	4.6	200					
9	28590683	72338	28590685	16	2	17	2	58	10	130	23	361	63	4.4	582					
10	1.34E+08	30671	1.34E+08	2	3	0	0	0	0	6	12	45	85	4.7	53					
11	1.1E+08	20556	1.1E+08	1	1	1	1	1	1	14	21	51	76	4.7	68					
12	2.14E+08	11395	2.14E+08	0	0	0	0	0	0	1	100	0	0	4	1					
13	2.01E+08	11395	2.01E+08	0	0	0	0	0	0	0	0	0	0	0	0					
14	1.05E+08	179250	1.05E+08	2	4	2	4	0	0	4	10	35	82	4.6	43					
15	2.72E+08	274962	2.72E+08	0	0	0	0	0	0	0	0	1	100	5	1					
16	1.77E+08	281904	1.77E+08	0	0	0	0	0	0	0	0	0	0	0	0					
17	92161380	179250	92161381	2	3	3	5	0	0	9	15	45	77	4.6	59					
18	32726343	72338	32726345	2	1	4	1	26	11	53	24	139	63	4.4	224					
19	1.04E+08	76045	1.04E+08	1	3	0	0	0	0	7	22	23	75	4.6	31					
20	20117907	108828	68229387	2	5	0	0	5	13	5	14	25	68	4.4	37					
21	86860057	9460	86860061	0	0	0	0	0	0	0	0	12	100	5	12					
22	2.55E+08	11395	2.55E+08	0	0	0	0	0	0	0	0	0	0	0	0					
23	2.53E+08	11395	2.53E+08	0	0	0	0	0	0	0	0	0	0	0	0					
24	2.53E+08	11395	2.53E+08	0	0	0	0	0	0	0	0	0	0	0	0					
25	77926784	9134	77926796	1	7	0	0	0	0	1	8	11	85	4.6	13					
26	92161380	179250	92161381	2	3	3	5	0	0	9	15	45	77	4.6	59					
27	70175132	104276	70175134	13	4	5	1	17	5	71	22	219	68	4.5	325					
28	63145867	11871	63145879	1	1	0	0	1	1	12	11	90	87	4.8	104					
29	1.97E+08	9460	1.97E+08	0	0	0	0	2	4	8	16	40	80	4.8	50					
30	31583583	72338	31583587	6	4	5	3	14	10	28	22	84	61	4.3	138					

giay_dep_nu_stars

Ready | Accessibility: Unavailable | 100%

