

```
In [7]: import pandas as pd
import numpy as np
import requests
from bs4 import BeautifulSoup
from urllib import request
```

```
In [8]: #Posting url in the web scraper tool
url = "https://www.jumia.co.ke/all-products/"
requests = requests.get(url)
```

```
In [9]: ## Parsing the html soup
soup = BeautifulSoup(requests.text, "html.parser")
```

```
In [10]: #searching for elements
items = soup.find_all("article", attrs = {"class": "prd _fb col c-prd"})
```

```
In [11]: print(len(items))
```

40

```
In [12]: print(items[0])
```

```
<article class="prd_fb col c-prd"><a class="btn_i_rnd -mas -fsh0 -me-start_wsl
t_sec" data-pop-open="addToWishlist" data-pop-trig="atw" data-simplesku="GE840HA2
62E1BNAFAMZ-153555440" data-sku="GE840HA262E1BNAFAMZ" data-track-onclick="wishlis
t" href="/customer/account/login/?tkwl=GE840HA262E1BNAFAMZ-153555440&return=%2
Fall-products%2F" rel="nofollow"><svg aria-label="Add to wishlist" class="ic -f-or
5" height="16" viewBox="0 0 24 24" width="16"><use xlink:href="https://www.jumia.c
o.ke/assets_he/images/i-icons.a66628fd.svg#saved-items"></use></svg></a><a class
="core" data-ga4-discount="14.09" data-ga4-item_brand="Generic" data-ga4-item_cate
gory="Home & Office" data-ga4-item_category2="Home & Kitchen" data-ga4-ite
m_category3="Kitchen & Dining" data-ga4-item_category4="Kitchen Utensils &
Gadgets" data-ga4-item_category5="Measuring Tools & Scales" data-ga4-item_cate
gory6="Scales" data-ga4-item_category7="Digital Scales" data-ga4-item_id="GE840HA2
62E1BNAFAMZ" data-ga4-item_name="Crane" data-ga4-list="" data-ga4-position="1" dat
a-ga4-price="45.81" data-gtm-brand="Generic" data-gtm-category="Home & Office/
Home & Kitchen/Kitchen & Dining/Kitchen Utensils & Gadgets/Measuring T
ools & Scales/Scales/Digital Scales" data-gtm-dimension23="109138" data-gtm-di
mension26="" data-gtm-dimension27="" data-gtm-dimension28="0" data-gtm-dimension37
="0" data-gtm-dimension43="" data-gtm-dimension44="0" data-gtm-id="GE840HA262E1BNA
FAMZ" data-gtm-list="" data-gtm-name="Crane" data-gtm-position="1" data-gtm-price
="45.81" data-track-onclick="eecProduct" data-track-onview="eecProduct" href="/gen
eric-crane-115512113.html"><div class="img-c"></div><div class="info"><h3
class="name">Crane</h3><div class="prc">KSh 6,500</div><div class="s-prc-w"><div c
lass="old">KSh 8,500</div><div class="bdg_dsct_sm">24%</div></div></div></a><foo
ter class="ft"><form action="/cart/products/GE840HA262E1BNAFAMZ-153555440/quantit
y/" class="-df -i-ctr -pr -j-bet" data-add-cart="" data-ga4-discount="14.09" data-
ga4-item_brand="Generic" data-ga4-item_category="Home & Office" data-ga4-item_
category2="Home & Kitchen" data-ga4-item_category3="Kitchen & Dining" data
-ga4-item_category4="Kitchen Utensils & Gadgets" data-ga4-item_category5="Meas
uring Tools & Scales" data-ga4-item_category6="Scales" data-ga4-item_category7
="Digital Scales" data-ga4-item_id="GE840HA262E1BNAFAMZ" data-ga4-item_name="Cran
e" data-ga4-price="45.81" data-gtm-brand="Generic" data-gtm-category="Home & O
ffice/Home & Kitchen/Kitchen & Dining/Kitchen Utensils & Gadgets/Measu
ring Tools & Scales/Scales/Digital Scales" data-gtm-dimension23="109138" data-
gtm-dimension26="" data-gtm-dimension27="" data-gtm-dimension28="0" data-gtm-dimen
sion37="0" data-gtm-dimension43="" data-gtm-dimension44="0" data-gtm-id="GE840HA26
2E1BNAFAMZ" data-gtm-name="Crane" data-gtm-price="45.81" data-gtm-simplesku="GE840
HA262E1BNAFAMZ-153555440" data-gtm-variant="" data-sku="GE840HA262E1BNAFAMZ" data-
svar="GE840HA262E1BNAFAMZ-153555440" data-track-onsubmit="addToCart" data-xhr="tru
e" method="post"><button class="add btn_prim -pea_md" data-submit="" name="actio
n" type="button" value="in">Add To Cart</button><input name="capiId" type="hidden"
value="_1714498507_{eventType}"/><input name="csrfToken" type="hidden" value="de29
74e91a52e2b844c30d8b44e2dc32"/></form></footer></article>
```

```
In [13]: first_items = items[0]
first_items_name = first_items.find("h3", attrs = {"class": "name"}).text
```

```
In [14]: print(first_items_name)
```

Crane

```
In [15]: ##price
first_items = items[0]
first_items_price = first_items.find("div", attrs = {"class": "prc"}).text
```

```
In [16]: print(first_items_price)
```

KSh 6,500

```
In [17]: #old price
first_items = items[0]
```

```
first_items_old_price = first_items.find("div", attrs = {"class": "old"})
```

```
In [18]: print(first_items_old_price)

<div class="old">KSh 8,500</div>
```

```
In [14]: ##discount
first_items = items[0]
first_items_discount = first_items.find("div", attrs = {"class": "bdg_dsct_sm"})
```

```
In [15]: print(first_items_discount)

<div class="bdg_dsct_sm">13%</div>
```

```
In [16]: ##product link
first_items = items[0]
first_items_link = first_items.find("a", attrs = {"class": "core"}).get("href")
```

```
In [17]: print(first_items_link)

/bluetooth-portable-speaker-with-fmus-b-c15-mpg1795907.html
```

```
In [18]: records = []
for item in items:
    try:
        item_name = item.find("h3", attrs={"class": "name"}).text
        current_price = item.find("div", attrs={"class": "prc"}).text
        old_price_element = item.find("div", attrs={"class": "old"})
        item_discount_element = item.find("div", attrs={"class": "bdg_dsct_sm"})
        item_link = item.find("a", attrs={"class": "core"}).get("href")

        if old_price_element:
            old_price = old_price_element.text
        else:
            old_price = None
        if item_discount_element:
            item_discount = item_discount_element.text
        else:
            item_discount = None
        records.append((item_name, current_price, old_price, item_discount, item_li

    except AttributeError:
        records.append(None)

print(len(records))
```

40

```
In [19]: records
```

```

Out[19]: [('C15 BASS Bluetooth Portable Speaker With FM/USB',
          'KSh 1,386',
          'KSh 1,600',
          '13%',
          '/bluetooth-portable-speaker-with-fmusbc15-mpg1795907.html'),
          ('Amtec Sub Woofer Sound System-BT/USB/FM-8,000 WATTS-',
          'KSh 4,639',
          'KSh 5,500',
          '16%',
          '/sub-woofer-sound-system-btusb-fm-8000-watts-amtec-mpg576273.html'),
          ('Istanbul Modern TV Stand With LED Lights',
          'KSh 9,999',
          'KSh 14,000',
          '29%',
          '/generic-istanbul-modern-tv-stand-with-led-lights-111914052.html'),
          ('Vitron 43" Inch Smart Android Bluetooth Connectivity Television, Inbuilt Deco
          der Netflix Youtube VGA HDMI USB+FREE BRACKET',
          'KSh 22,899',
          'KSh 34,999',
          '35%',
          '/vitron-43-inch-smart-android-bluetooth-connectivity-television-inbuilt-decoder
          -netflix-youtube-vga-hdmi-usb-free-bracket-191428609.html'),
          ('C Idea Adults Tablets, 10 Inches Android 12 Sim 8GB+512GB Storage Keyboard (Re
          d)',
          'KSh 13,504',
          'KSh 21,134',
          '36%',
          '/c-idea-adults-tablets-10-inches-android-12-sim-8gb512gb-storage-keyboard-red-1
          60581068.html'),
          ('Wireless Bluetooth Headset Gaming Earphones Led Dispaly ',
          'KSh 650',
          'KSh 1,000',
          '35%',
          '/generic-wireless-bluetooth-headset-gaming-earphones-led-dispaly-169311275.htm
          l'),
          ('Vitron V636, 3.1CH Multimedia Bluetooth Speaker System - Black',
          'KSh 5,670',
          'KSh 8,059',
          '30%',
          '/vitron-v636-3.1ch-multimedia-bluetooth-speaker-system-black-25652234.html'),
          ('Full Touch Smart Watch Sport Health BT Call',
          'KSh 1,220',
          'KSh 1,930',
          '37%',
          '/generic-full-touch-smart-watch-sport-health-bt-call-152070136.html'),
          ('MK TV Guard - Surge Protector, + FREE Gift',
          'KSh 650',
          'KSh 700',
          '7%',
          '/mk-tv-guard-surge-protector-free-gift-47911215.html'),
          ('Tecno T301; (Dual Sim),// Memory Card Slot//black',
          'KSh 1,790',
          'KSh 2,000',
          '11%',
          '/t301-dual-sim-memory-card-slotblack-tecno-mpg1568337.html'),
          ('Solarmax 100W Solar Panel Full Kit All Weather+ 80AH Solar Battery + 300 Watts
          Solar Inverter + 10AH Controller + 5 DC Bulbs + 10M Cable 25 YEARS WARRANT',
          'KSh 12,900',
          'KSh 24,000',
          '46%',
          '/solarmax-100w-solar-panel-full-kit-all-weather-80ah-solar-battery-300-watts-so
          lar-inverter-10ah-controller-5-dc-bulbs-10m-cable-25-years-warrant-133071709.htm
          l'),
          ('XIAOMI Mi 300Mbps Wifi Range Extender Repeater Pro Amplifier',

```

```

'KSh 2,200',
'KSh 3,000',
'27%',
'/xiaomi-mi-300mbps-wifi-range-extender-repeater-pro-amplifier-125031457.html'),
('Amaya HD-360BT Active Noise Cancellation (ANC) Over-Ear Wireless Bluetooth Headphones.',
'KSh 4,999',
'KSh 5,500',
'9%',
'/amaya-hd-360bt-active-noise-cancellation-anc-over-ear-wireless-bluetooth-headphones.-210744502.html'),
('F9 Bluetooth Earphone 2000mAh Battery Headset Zw',
'KSh 490',
'KSh 799',
'39%',
'/generic-f9-bluetooth-earphone-2000mah-battery-headset-zw-58945935.html'),
('Bontel T1000, Wireless Desktop Phone, Sms,,Feature- Black,,home And Office',
'KSh 2,920',
'KSh 4,500',
'35%',
'/t1000-wireless-desktop-phone-smsfeature-blackhome-and-office-bontel-mpg1568309.html'),
('DELL Wired Laser Mouse',
'KSh 379',
'KSh 499',
'24%',
'/dell-wired-laser-mouse-152293841.html'),
('Tecno Spark 20, 6.6'' HD+, UP to 8GB RAM + 128GB ROM (Dual Sim) 50MP, 5000 mAh - Black",
'KSh 14,700',
'KSh 14,750',
'1%',
'/tecno-spark-20-6.6-hd-up-to-8gb-ram-128gb-rom-dual-sim-50mp-5000-mah-black-203139356.html'),
('Freeyond Smartwatch Sports IP67 Running FY Watch F1',
'KSh 1,520',
'KSh 2,881',
'47%',
'/freeyond-smartwatch-sports-ip67-running-fy-watch-f1-126637609.html'),
('Phone Vlog Video Fill Light Kit With Adjustable Tripod Mic',
'KSh 2,970',
'KSh 3,989',
'26%',
'/generic-phone-vlog-video-fill-light-kit-with-adjustable-tripod-mic-152070161.html'),
('Fashion Basketball Shoes Wear-resistant Men's Basketball Shoes",
'KSh 1,499',
'KSh 2,599',
'42%',
'/fashion-basketball-shoes-wear-resistant-mens-basketball-shoes-166897705.html'),
('Headphone Microphone Audio Splitter Adapter',
'KSh 299',
'KSh 498',
'40%',
'/generic-headphone-microphone-audio-splitter-adapter-167158931.html'),
('Apple 35W Original Fast Charger For iPhone 11/12/13/ 14/14Pro/14 Max/ 14 Pro Max',
'KSh 2,999',
None,
None,
'/apple-35w-original-fast-charger-for-iphone-111213-1414pro14-max-14-pro-max-159766037.html'),
('Vitron HTC4388FS,43'' SMART TV Android Frameless TV,NetflixYoutube+FREE TVGuard+
```

```

Bracket",
  'KSh 22,495',
  'KSh 30,000',
  '25%',
  '/vitron-htc4388fs43smart-tv-android-frameless-tvnetflixyoutube-free-tvguardbracket-30832336.html'),
  ('Outdoor/Indoor HDTV Digital TV Antenna, Signal Booster, Range Up To 150 Miles',
  'KSh 1,880',
  'KSh 3,080',
  '39%',
  '/generic-outdoorindoor-hdtv-digital-tv-antenna-signal-booster-range-up-to-150-miles-156383024.html'),
  ('Darkhorsir Fast Built-in 4Cables Real 20,000 MAh Power Bank B',
  'KSh 2,199',
  'KSh 3,250',
  '32%',
  '/darkhorsir-fast-built-in-4cables-real-20000-mah-power-bank-b-49111521.html'),
  ('5V 3A Charger Head Digital Display High Compatibility-White',
  'KSh 379',
  'KSh 758',
  '50%',
  '/generic-5v-3a-charger-head-digital-display-high-compatibility-white-67769534.html'),
  ('HDMI CABLES 1.5M ,3M ,5M ,10M ,15M ,20M ,30M',
  'KSh 200 - KSh 2,600',
  'KSh 300 - KSh 2,800',
  '33%',
  '/generic-hdmi-cables-1.5m-3m-5m-10m-15m-20m-30m-167946703.html'),
  ('Tripod Suitable For Camera Phone Holder',
  'KSh 1,199',
  'KSh 2,099',
  '43%',
  '/generic-tripod-suitable-for-camera-phone-holder-176767482.html'),
  ('X Tigi HeroBook 10.1"-Quad Core-4Gb RAM-128Gb ROM -Android Notebook- 4000mAh-13 Months Warranty',
  'KSh 11,999',
  'KSh 21,599',
  '44%',
  '/x-tigi-herobook-10.1-quad-core-4gb-ram-128gb-rom-android-notebook-4000mah-13-months-warranty-106952479.html'),
  ('Wireless Bluetooth Earbuds Earphone HIFI Stereo Pods3 Air',
  'KSh 529',
  'KSh 998',
  '47%',
  '/wireless-bluetooth-earphone-earbuds-hifi-stereo-touch-sensor-control-for-samsung-iphone-apple-xiaomi-huawei-vivo-iosandroid-generic-mpg358778.html'),
  ('Y1 Bluetooth SPORT Smartwatch GSM SIM SD CARD SLOT',
  'KSh 1,899',
  'KSh 2,000',
  '5%',
  '/generic-y1-bluetooth-sport-smartwatch-gsm-sim-sd-card-slot-125902631.html'),
  ('Vitron 32- Inch Frameless Smart Android TV,Netflix,Facebook,Playstore+BRACKET+GUARD',
  'KSh 14,599',
  'KSh 26,000',
  '44%',
  '/32-inch-frameless-smart-android-tvnetflixfacebookplaystorebracketguard-vitron-mpg1662660.html'),
  ('Vitron Frameless 32" Inch Smart Android TV Netflix-Appstore+FREE BRACKET',
  'KSh 13,699',
  'KSh 26,000',
  '47%',
  '/htc3200s32inch-framelss-television-smart-android-tv-freebracket-vitron-mpg1522006.html'),

```

```
( 'Car Modulator Wireless FM Adapter With Bluetooth',
  'KSh 543',
  'KSh 1,000',
  '46%',
  '/generic-car-modulator-wireless-fm-adapter-with-bluetooth-43686451.html'),
( 'Baofeng BF-888S Walkie Talkie Two Way Handheld Radio- 8PCS',
  'KSh 10,999',
  'KSh 15,999',
  '31%',
  '/baofeng-bf-888s-walkie-talkie-two-way-handheld-radio-8pcs-159247737.html'),
( 'Kensplastic Knapsack Sprayer/Backpack Sprayer 16lt',
  'KSh 2,200',
  'KSh 2,800',
  '21%',
  '/kensplastic-knapsack-sprayerbackpack-sprayer-16lt-58706192.html'),
( 'BVOT Universal 4G/5G Portable Wireless Wi-fi- Recommended',
  'KSh 4,599',
  'KSh 6,999',
  '34%',
  '/bvot-universal-4g5g-portable-wireless-wi-fi-recommended-200300689.html'),
( 'Advance High performance Micro SD HC Memory Card - 64gb plus free gift ',
  'KSh 549',
  None,
  None,
  '/advance-high-performance-micro-sd-hc-memory-card-64gb-plus-free-gift-191247645.html'),
( 'Nobel 5.1CH Super Woofer Speaker Fm/usb/bt',
  'KSh 13,680',
  'KSh 17,000',
  '20%',
  '/nobel-5.1ch-super-woofer-speaker-fmusbbt-179315519.html'),
( 'Mouse Pad Relieve Stress Comfy EVA Wrist Support-Black',
  'KSh 195',
  'KSh 293',
  '33%',
  '/generic-mouse-pad-relieve-stress-comfy-eva-wrist-support-black-77714730.html')]
```

```
In [20]: import time
import requests
from bs4 import BeautifulSoup

base_url = "https://www.jumia.co.ke/all-products/?page={}&catalog-listing"
records = []

#for loop is used for loop all 50 pages of jumia website

for page_number in range(1, 51):
    url = base_url.format(page_number)
    time.sleep(5)
    response = requests.get(url)
    if response.status_code == 200:
        soup = BeautifulSoup(response.text, "html.parser")
        print(f"Page {page_number} scrapped successfully")
        items = soup.find_all("article", attrs={"class": "prd_fb col c-prd"})
        for item in items:
            try:
                item_name = item.find("h3", attrs={"class": "name"}).text
                current_price = item.find("div", attrs={"class": "prc"}).text
                old_price_element = item.find("div", attrs={"class": "old"})
                item_discount_element = item.find("div", attrs={"class": "bdg_dsct"})
                item_link = item.find("a", attrs={"class": "core"}).get("href")

                if old_price_element:
```

```
        old_price = old_price_element.text
    else:
        old_price = None
    if item_discount_element:
        item_discount = item_discount_element.text
    else:
        item_discount = None
    records.append((item_name, current_price, old_price, item_discount,
except AttributeError:
    records.append(None)
else:
    print(f"Page {page_number} not scrapped")

# Process records further as needed
```


Page 1 scrapped successfully
Page 2 scrapped successfully
Page 3 scrapped successfully
Page 4 scrapped successfully
Page 5 scrapped successfully
Page 6 scrapped successfully
Page 7 scrapped successfully
Page 8 scrapped successfully
Page 9 scrapped successfully
Page 10 scrapped successfully
Page 11 scrapped successfully
Page 12 scrapped successfully
Page 13 scrapped successfully
Page 14 scrapped successfully
Page 15 scrapped successfully
Page 16 scrapped successfully
Page 17 scrapped successfully
Page 18 scrapped successfully
Page 19 scrapped successfully
Page 20 scrapped successfully
Page 21 scrapped successfully
Page 22 scrapped successfully
Page 23 scrapped successfully
Page 24 scrapped successfully
Page 25 scrapped successfully
Page 26 scrapped successfully
Page 27 scrapped successfully
Page 28 scrapped successfully
Page 29 scrapped successfully
Page 30 scrapped successfully
Page 31 scrapped successfully
Page 32 scrapped successfully
Page 33 scrapped successfully
Page 34 scrapped successfully
Page 35 scrapped successfully
Page 36 scrapped successfully
Page 37 scrapped successfully
Page 38 scrapped successfully
Page 39 scrapped successfully
Page 40 scrapped successfully
Page 41 scrapped successfully
Page 42 scrapped successfully
Page 43 scrapped successfully
Page 44 scrapped successfully
Page 45 scrapped successfully
Page 46 scrapped successfully
Page 47 scrapped successfully
Page 48 scrapped successfully
Page 49 scrapped successfully
Page 50 scrapped successfully

In [21]: `len(records)`

Out[21]: 2000

In [22]: `#converting list to dataframe with the help of pandas library`
`data = pd.DataFrame(records, columns = ["item_name", "current_price", "old_price_eleme`

In [23]: `data.head()`

Out[23]:

	item_name	current_price	old_price_element	item_discount_element	item_link
0	Fashion Men's Shoes Formal Casual Shoes Shoes ...	KSh 1,195	KSh 1,991	40%	/fashion-mens-shoes-formal-casual-shoes-shoes-...
1	Fashion Mens Sneaker Causal Breathable Shoes L...	KSh 1,000	KSh 2,000	50%	/fashion-mens-sneaker-causal-breathable-shoes-...
2	Fashion Men Shoes Loafers Casual Shoes Slip-On...	KSh 699	KSh 1,156	40%	/fashion-mens-shoes-loafers-casual-shoes-slip-o...
3	Fashion Men's Sneakers Breathable Sports Shoe...	KSh 699	KSh 1,000	30%	/fashion-mens-sneakers-breathable-sports-shoes...
4	Fashion 2024 Mens Casual Air-Cushion Shoes Run...	KSh 1,490	KSh 2,223 - KSh 2,865	48%	/fashion-2024-mens-casual-air-cushion-shoes-ru...

In [24]: data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   item_name              2000 non-null   object
1   current_price          2000 non-null   object
2   old_price_element      1941 non-null   object
3   item_discount_element  1941 non-null   object
4   item_link              2000 non-null   object
dtypes: object(5)
memory usage: 78.3+ KB
```

```
In [25]: ## filtering out prices that have a range
data['current_price'] = data['current_price'].apply(lambda x: x[4:] )
```

```
In [26]: data = data.dropna()
```

```
In [27]: data['old_price_element'] = data['old_price_element'].apply(lambda x: x[4:] )
data['item_discount_element'] = data['item_discount_element'].apply(lambda x: x[:2]
data
```

Out[27]:

	item_name	current_price	old_price_element	item_discount_element	item_lir
0	Fashion Men's Shoes Formal Casual Shoes Shoes ...	1,195	1,991	40	/fashion-men-shoes-forma-casual-shoe-shoes-
1	Fashion Mens Sneaker Causal Breathable Shoes L...	1,000	2,000	50	/fashion-men-sneaker-causa-breathabl-shoes-
2	Fashion Men Shoes Loafers Casual Shoes Slip-On...	699	1,156	40	/fashion-mei-shoes-loafer-casual-shoes-slij-c
3	Fashion Men's Sneakers Breathable Sports Shoe...	699	1,000	30	/fashion-men-sneaker-breathable-sport-shoes-
4	Fashion 2024 Mens Casual Air-Cushion Shoes Run...	1,490	2,223 - KSh 2,865	48	/fashion-202-mens-casual-ai-cushion-shoes-ru
...	
1995	Dangling Tassels Shoulder Chain Necklace Body ...	2,220	3,559	38	/generic-danglin-tassels-shoulder-chain-neck
1996	Skinny Pin Barbell Gentleman Collar Tie Clip C...	592	984	40	/generic-skinny-pin-barbell-gentleman-colla-t
1997	100*155cm Artificial Leather Fabric For Upy Fu...	7,178	12,284	42	/generi-100155cm-artificial-leathe-fabric-fc
1998	Posture correctors/supports for men and women ...	1,021	1,524	33	/generic-postur-correctorssupport-for-men-ar
1999	1080P Digital Camera 20MP Kids Camera Cute Chi...	1,980	3,760	47	/generic-1080p-digital-camer-20mp-kid-camera

1941 rows × 5 columns



In [28]:

```
#current price contains starting price and end price so we have to split this column
selected_rows = data[data['current_price'].str.contains('-') | data['old_price_element'].str.contains('-')]
selected_rows
```

Out[28]:

	item_name	current_price	old_price_element	item_discount_element	item_link
4	Fashion 2024 Mens Casual Air-Cushion Shoes Run...	1,490	2,223 - KSh 2,865	48	/fashion-2024-mens-casual-air-cushion-shoes-ru...
6	Fashion Men's Casual Sneakers Black	1,290 - KSh 1,300	2,027 - KSh 2,206	42	/fashion-mens-casual-sneakers-black-58273521.html
25	Fashion Boys Shoes Girls Shoes Children's Canv...	559 - KSh 600	1,200	53	/fashion-boys-shoes-girls-shoes-childrens-canv...
31	Fashion 4 Pcs Men Khaki Trousers Stretch Slim Fit	3,980 - KSh 4,900	4,500 - KSh 5,500	13	/fashion-4-pcs-men-khaki-trousers-stretch-slim-...
32	Fashion 3 Pure Cotton Men's & Husband's Boxers...	1,300 - KSh 1,480	2,190	41	/fashion-3-pure-cotton-mens-husbands-boxers-un...
33	Fashion Mens Cotton Zipper Jacket Autumn&Winte...	650 - KSh 1,350	1,275 - KSh 2,500	50	/fashion-mens-cotton-zipper-jacket-autumnwinte...
38	Fashion VERSATILE Men's PURE Leather Loafer S...	1,239 - KSh 1,299	1,599	23	/fashion-versatile-mens-pure-leather-loafer-sh...
259	Reusable Plant Cover Adjustable Drawstring Mes...	1,622 - KSh 2,427	2,433 - KSh 3,641	33	/generic-reusable-plant-cover-adjustable-draws...
974	5 Speed/ 6 Gear Car Shift Gear change lever kn...	5,180 - KSh 5,328	5,920 - KSh 6,068	13	/generic-5-speed-6-gear-car-shift-gear-change-...
1347	Baby Boys Suit Set Toddler Blazer Infant Weddi...	4,896 - KSh 5,328	8,296 - KSh 9,028	41	/generic-baby-boys-suit-set-toddler-blazer-inf...
1723	Women's Knit Lightweight Mesh Sneakers, Breath...	6,538 - KSh 6,780	12,422 - KSh 12,881	47	/generic-womens-knit-lightweight-mesh-sneakers...
1939	Hot T16 Wireless Bluetooth Headset 5.3 Led Dig...	4,144 - KSh 4,884	5,624 - KSh 6,364	26	/generic-hot-t16-wireless-bluetooth-headset-5....

	item_name	current_price	old_price_element	item_discount_element	item_link
1981	Cartoon Toilet Seat Cover Pad Summer Four Seas...	3,402 - KSh 3,564	5,022 - KSh 5,184	32	/generic- cartoon-toilet- seat-cover- pad- summer-...

split the columns with price ranges to only remain with one price value

```
In [29]: #begin with the current price columns
split_values = selected_rows['current_price'].str.split('-', expand=True)

# Extract start_value and end_value from split_values
selected_rows.loc[:, 'price_start_value'] = split_values[0].str.strip()
selected_rows.loc[:, 'price_end_value'] = split_values[1].str.strip()

# Fill missing end_value with start_value
selected_rows['price_end_value'].fillna(selected_rows['price_start_value'], inplace=True)

selected_rows
```

C:\Users\devra\AppData\Local\Temp\ipykernel_12164\3716473462.py:6: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
selected_rows.loc[:, 'price_start_value'] = split_values[0].str.strip()
```

C:\Users\devra\AppData\Local\Temp\ipykernel_12164\3716473462.py:7: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
selected_rows.loc[:, 'price_end_value'] = split_values[1].str.strip()
```

C:\Users\devra\AppData\Local\Temp\ipykernel_12164\3716473462.py:10: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
selected_rows['price_end_value'].fillna(selected_rows['price_start_value'], inplace=True)
```

Out[29]:

	item_name	current_price	old_price_element	item_discount_element	item_link	price
4	Fashion 2024 Mens Casual Air-Cushion Shoes Run...	1,490	2,223 - KSh 2,865	48	/fashion-2024-mens-casual-air-cushion-shoes-ru...	
6	Fashion Men's Casual Sneakers Black	1,290 - KSh 1,300	2,027 - KSh 2,206	42	/fashion-mens-casual-sneakers-black-58273521.html	
25	Fashion Boys Shoes Girls Shoes Children's Canv...	559 - KSh 600	1,200	53	/fashion-boys-shoes-girls-shoes-childrens-canv...	
31	Fashion 4 Pcs Men Khaki Trousers Stretch Slim Fit	3,980 - KSh 4,900	4,500 - KSh 5,500	13	/fashion-4-pcs-men-khaki-trousers-stretch-slim-...	
32	Fashion 3 Pure Cotton Men's & Husband's Boxers...	1,300 - KSh 1,480	2,190	41	/fashion-3-pure-cotton-mens-husbands-boxers-un...	
33	Fashion Mens Cotton Zipper Jacket Autumn&Winte...	650 - KSh 1,350	1,275 - KSh 2,500	50	/fashion-mens-cotton-zipper-jacket-autumnwinte...	
38	Fashion VERSATILE Men's PURE Leather Loafer S...	1,239 - KSh 1,299	1,599	23	/fashion-versatile-mens-pure-leather-loafer-sh...	
259	Reusable Plant Cover Adjustable Drawstring Mes...	1,622 - KSh 2,427	2,433 - KSh 3,641	33	/generic-reusable-plant-cover-adjustable-draws...	
974	5 Speed/ 6 Gear Car Shift Gear change lever kn...	5,180 - KSh 5,328	5,920 - KSh 6,068	13	/generic-5-speed-6-gear-car-shift-gear-change-...	
1347	Baby Boys Suit Set Toddler Blazer Infant Weddi...	4,896 - KSh 5,328	8,296 - KSh 9,028	41	/generic-baby-boys-suit-set-toddler-blazer-inf...	
1723	Women's Knit Lightweight Mesh Sneakers, Breath...	6,538 - KSh 6,780	12,422 - KSh 12,881	47	/generic-womens-knit-lightweight-mesh-sneakers...	
1939	Hot T16 Wireless Bluetooth	4,144 - KSh 4,884	5,624 - KSh 6,364	26	/generic-hot-t16-wireless-	

	item_name	current_price	old_price_element	item_discount_element	item_link	price
	Headset 5.3 Led Dig...				bluetooth- headset-5....	
1981	Cartoon Toilet Seat Cover Pad Summer Four Seas...	3,402 - KSh 3,564	5,022 - KSh 5,184	32	/generic- cartoon-toilet- seat-cover- pad-	

```
In [30]: ##old price column
split_values = selected_rows['old_price_element'].str.split('-', expand=True)

# Extract start_value and end_value from split_values
selected_rows.loc[:, 'old_price_start_value'] = split_values[0].str.strip()
selected_rows.loc[:, 'old_price_end_value'] = split_values[1].str.strip()

# Fill missing end_value with start_value
selected_rows['old_price_end_value'].fillna(selected_rows['old_price_start_value'],
selected_rows

C:\Users\devra\AppData\Local\Temp\ipykernel_12164\3216186134.py:5: SettingWithCopy
Warning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stabl
e/user_guide/indexing.html#returning-a-view-versus-a-copy
selected_rows.loc[:, 'old_price_start_value'] = split_values[0].str.strip()
C:\Users\devra\AppData\Local\Temp\ipykernel_12164\3216186134.py:6: SettingWithCopy
Warning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stabl
e/user_guide/indexing.html#returning-a-view-versus-a-copy
selected_rows.loc[:, 'old_price_end_value'] = split_values[1].str.strip()
C:\Users\devra\AppData\Local\Temp\ipykernel_12164\3216186134.py:9: SettingWithCopy
Warning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stabl
e/user_guide/indexing.html#returning-a-view-versus-a-copy
selected_rows['old_price_end_value'].fillna(selected_rows['old_price_start_valu
e'], inplace=True)
```

Out[30]:

	item_name	current_price	old_price_element	item_discount_element	item_link	price
4	Fashion 2024 Mens Casual Air-Cushion Shoes Run...	1,490	2,223 - KSh 2,865	48	/fashion-2024-mens-casual-air-cushion-shoes-ru...	
6	Fashion Men's Casual Sneakers Black	1,290 - KSh 1,300	2,027 - KSh 2,206	42	/fashion-mens-casual-sneakers-black-58273521.html	
25	Fashion Boys Shoes Girls Shoes Children's Canv...	559 - KSh 600	1,200	53	/fashion-boys-shoes-girls-shoes-childrens-canv...	
31	Fashion 4 Pcs Men Khaki Trousers Stretch Slim Fit	3,980 - KSh 4,900	4,500 - KSh 5,500	13	/fashion-4-pcs-men-khaki-trousers-stretch-slim-...	
32	Fashion 3 Pure Cotton Men's & Husband's Boxers...	1,300 - KSh 1,480	2,190	41	/fashion-3-pure-cotton-mens-husbands-boxers-un...	
33	Fashion Mens Cotton Zipper Jacket Autumn&Winte...	650 - KSh 1,350	1,275 - KSh 2,500	50	/fashion-mens-cotton-zipper-jacket-autumnwinte...	
38	Fashion VERSATILE Men's PURE Leather Loafer S...	1,239 - KSh 1,299	1,599	23	/fashion-versatile-mens-pure-leather-loafer-sh...	
259	Reusable Plant Cover Adjustable Drawstring Mes...	1,622 - KSh 2,427	2,433 - KSh 3,641	33	/generic-reusable-plant-cover-adjustable-draws...	
974	5 Speed/ 6 Gear Car Shift Gear change lever kn...	5,180 - KSh 5,328	5,920 - KSh 6,068	13	/generic-5-speed-6-gear-car-shift-gear-change-...	
1347	Baby Boys Suit Set Toddler Blazer Infant Weddi...	4,896 - KSh 5,328	8,296 - KSh 9,028	41	/generic-baby-boys-suit-set-toddler-blazer-inf...	
1723	Women's Knit Lightweight Mesh Sneakers, Breath...	6,538 - KSh 6,780	12,422 - KSh 12,881	47	/generic-womens-knit-lightweight-mesh-sneakers...	
1939	Hot T16 Wireless Bluetooth	4,144 - KSh 4,884	5,624 - KSh 6,364	26	/generic-hot-t16-wireless-	

	item_name	current_price	old_price_element	item_discount_element	item_link	price
	Headset 5.3 Led Dig...				bluetooth- headset-5...	
1981	Cartoon Toilet Seat Cover Pad Summer Four Seas...	3,402 - KSh 3,564	5,022 - KSh 5,184	32	/generic- cartoon-toilet- seat-cover- pad-	

In [31]: `selected_rows.drop(['current_price', 'old_price_element'], axis = 1, inplace=True)`

C:\Users\devra\AppData\Local\Temp\ipykernel_12164\1709597164.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

`selected_rows.drop(['current_price', 'old_price_element'], axis = 1, inplace=True)`

In [32]: `columns_to_convert = ['price_start_value', 'price_end_value', 'old_price_start_value', 'old_price_end_value']`

Function to convert string values to float, handling 'KSh' and commas

def convert_to_float(value):

if isinstance(value, str):

if 'KSh' in value:

 value = value.replace('KSh', '').replace(',', '').strip()

else:

 value = value.replace(',', '').strip()

return pd.to_numeric(value, errors='coerce')

Convert values in specified columns to float

for column **in** columns_to_convert:

 selected_rows[column] = selected_rows[column].apply(convert_to_float)

selected_rows

C:\Users\devra\AppData\Local\Temp\ipykernel_12164\1611824524.py:14: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

`selected_rows[column] = selected_rows[column].apply(convert_to_float)`

Out[32]:

	item_name	item_discount_element	item_link	price_start_value	price_end_value	ol
4	Fashion 2024 Mens Casual Air-Cushion Shoes Run...	48	/fashion-2024-mens-casual-air-cushion-shoes-ru...	1490	1490	
6	Fashion Men's Casual Sneakers Black	42	/fashion-mens-casual-sneakers-black-58273521.html	1290	1300	
25	Fashion Boys Shoes Girls Shoes Children's Canv...	53	/fashion-boys-shoes-girls-shoes-childrens-canv...	559	600	
31	Fashion 4 Pcs Men Khaki Trousers Stretch Slim Fit	13	/fashion-4-pcs-men-khaki-trousers-stretch-slim-...	3980	4900	
32	Fashion 3 Pure Cotton Men's & Husband's Boxers...	41	/fashion-3-pure-cotton-mens-husbands-boxers-un...	1300	1480	
33	Fashion Mens Cotton Zipper Jacket Autumn&Winte...	50	/fashion-mens-cotton-zipper-jacket-autumnwinte...	650	1350	
38	Fashion VERSATILE Men's PURE Leather Loafer S...	23	/fashion-versatile-mens-pure-leather-loafer-sh...	1239	1299	
259	Reusable Plant Cover Adjustable Drawstring Mes...	33	/generic-reusable-plant-cover-adjustable-draws...	1622	2427	
974	5 Speed/ 6 Gear Car Shift Gear change lever kn...	13	/generic-5-speed-6-gear-car-shift-gear-change-...	5180	5328	
1347	Baby Boys Suit Set Toddler Blazer Infant Weddi...	41	/generic-baby-boys-suit-set-toddler-blazer-inf...	4896	5328	
1723	Women's Knit Lightweight Mesh Sneakers, Breath...	47	/generic-womens-knit-lightweight-mesh-sneakers...	6538	6780	
1939	Hot T16 Wireless Bluetooth	26	/generic-hot-t16-wireless-	4144	4884	

	item_name	item_discount_element	item_link	price_start_value	price_end_value	ol
	Headset 5.3 Led Dig...		bluetooth-headset-5....			
1981	Cartoon Toilet Seat Cover Pad Summer Four Seas...	32	/generic-cartoon-toilet-seat-cover-pad-	3402	3564	

In [33]: `selected_rows['current_price']=(selected_rows['price_start_value'] + selected_rows['price_end_value'])/2
selected_rows['old_price_element']=(selected_rows['old_price_start_value'] + selected_rows['old_price_end_value'])/2
selected_rows.head(2)`

C:\Users\devra\AppData\Local\Temp\ipykernel_12164\2322384438.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

`selected_rows['current_price']=(selected_rows['price_start_value'] + selected_rows['price_end_value'])/2`

C:\Users\devra\AppData\Local\Temp\ipykernel_12164\2322384438.py:2: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

`selected_rows['old_price_element']=(selected_rows['old_price_start_value'] + selected_rows['old_price_end_value'])/2`

Out[33]:

	item_name	item_discount_element	item_link	price_start_value	price_end_value	old_price_s
4	Fashion 2024 Mens Casual Air-Cushion Shoes Run...	48	/fashion-2024-mens-casual-air-cushion-shoes-ru...	1490	1490	
6	Fashion Men's Casual Sneakers Black	42	/fashion-mens-casual-sneakers-black-58273521.html	1290	1300	

In [34]: `selected_rows.drop(['price_start_value', 'price_end_value', 'old_price_start_value', 'old_price_end_value'], axis=1, inplace=True)
selected_rows.head()`

C:\Users\devra\AppData\Local\Temp\ipykernel_12164\1953087594.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

`selected_rows.drop(['price_start_value', 'price_end_value', 'old_price_start_value', 'old_price_end_value'], axis=1, inplace=True)`

Out[34]:

	item_name	item_discount_element	item_link	current_price	old_price_element
4	Fashion 2024 Mens Casual Air-Cushion Shoes Run...	48	/fashion-2024-mens-casual-air-cushion-shoes-ru...	1490.0	2544.0
6	Fashion Men's Casual Sneakers Black	42	/fashion-mens-casual-sneakers-black-58273521.html	1295.0	2116.5
25	Fashion Boys Shoes Girls Shoes Children's Canv...	53	/fashion-boys-shoes-girls-shoes-childrens-canv...	579.5	1200.0
31	Fashion 4 Pcs Men Khaki Trouser Stretch Slim Fit	13	/fashion-4-pcs-men-khaki-trouser-stretch-slim-...	4440.0	5000.0
32	Fashion 3 Pure Cotton Men's & Husband's Boxers...	41	/fashion-3-pure-cotton-mens-husbands-boxers-un...	1390.0	2190.0

```
In [35]: selected_rows= selected_rows[['item_name','current_price','old_price_element','item_discount_element']]
selected_rows.head()
```

Out[35]:

	item_name	current_price	old_price_element	item_discount_element	item_link
4	Fashion 2024 Mens Casual Air-Cushion Shoes Run...	1490.0	2544.0	48	/fashion-2024-mens-casual-air-cushion-shoes-ru...
6	Fashion Men's Casual Sneakers Black	1295.0	2116.5	42	/fashion-mens-casual-sneakers-black-58273521.html
25	Fashion Boys Shoes Girls Shoes Children's Canv...	579.5	1200.0	53	/fashion-boys-shoes-girls-shoes-childrens-canv...
31	Fashion 4 Pcs Men Khaki Trouser Stretch Slim Fit	4440.0	5000.0	13	/fashion-4-pcs-men-khaki-trouser-stretch-slim-...
32	Fashion 3 Pure Cotton Men's & Husband's Boxers...	1390.0	2190.0	41	/fashion-3-pure-cotton-mens-husbands-boxers-un...

```
In [36]: new_df=data.drop(selected_rows.index)
new_df.head()
```

Out[36]:

	item_name	current_price	old_price_element	item_discount_element	item_link
0	Fashion Men's Shoes Formal Casual Shoes Shoes ...	1,195	1,991	40	/fashion-mens- shoes-formal- casual-shoes- shoes-...
1	Fashion Mens Sneaker Causal Breathable Shoes L...	1,000	2,000	50	/fashion-mens- sneaker-causal- breathable- shoes-...
2	Fashion Men Shoes Loafers Casual Shoes Slip-On...	699	1,156	40	/fashion-mens- shoes-loafers- casual-shoes-slip- o...
3	Fashion Men's Sneakers Breathable Sports Shoe...	699	1,000	30	/fashion-mens- sneakers- breathable- sports-shoes...
5	Fashion Men's Casual Shoes Sneakers Sports Shoes	1,199	1,299	8%	/fashion-mens- casual-shoes- sneakers-sports- sho...

In [37]:

```

columns_to_convert = ['current_price', 'old_price_element']

# Function to convert string values to float, handling 'KSh' and commas
def convert_to_float(value):
    if isinstance(value, str):
        if 'KSh' in value:
            value = value.replace('KSh', '').replace(',', '').strip()
        else:
            value = value.replace(',', '').strip()
    return pd.to_numeric(value, errors='coerce')

# Convert values in specified columns to float
for column in columns_to_convert:
    new_df[column] = new_df[column].apply(convert_to_float)
new_df.head()

```

Out[37]:

	item_name	current_price	old_price_element	item_discount_element	item_link
0	Fashion Men's Shoes Formal Casual Shoes Shoes ...	1195	1991	40	/fashion-mens-shoes-formal-casual-shoes-shoes-...
1	Fashion Mens Sneaker Causal Breathable Shoes L...	1000	2000	50	/fashion-mens-sneaker-causal-breathable-shoes-...
2	Fashion Men Shoes Loafers Casual Shoes Slip-On...	699	1156	40	/fashion-mens-shoes-loafers-casual-shoes-slip-o...
3	Fashion Men's Sneakers Breathable Sports Shoe...	699	1000	30	/fashion-mens-sneakers-breathable-sports-shoes-...
5	Fashion Men's Casual Shoes Sneakers Sports Shoes	1199	1299	8%	/fashion-mens-casual-shoes-sneakers-sports-sho...

```
In [38]: clean_df=pd.concat([new_df,selected_rows],ignore_index=True)
clean_df.head()
```

Out[38]:

	item_name	current_price	old_price_element	item_discount_element	item_link
0	Fashion Men's Shoes Formal Casual Shoes Shoes ...	1195.0	1991.0	40	/fashion-mens-shoes-formal-casual-shoes-shoes-...
1	Fashion Mens Sneaker Causal Breathable Shoes L...	1000.0	2000.0	50	/fashion-mens-sneaker-causal-breathable-shoes-...
2	Fashion Men Shoes Loafers Casual Shoes Slip-On...	699.0	1156.0	40	/fashion-mens-shoes-loafers-casual-shoes-slip-o...
3	Fashion Men's Sneakers Breathable Sports Shoe...	699.0	1000.0	30	/fashion-mens-sneakers-breathable-sports-shoes-...
4	Fashion Men's Casual Shoes Sneakers Sports Shoes	1199.0	1299.0	8%	/fashion-mens-casual-shoes-sneakers-sports-sho...

```
In [39]: clean_df['item_discount_element'] =clean_df['item_discount_element'].str.replace('%
```

```
In [40]: clean_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1941 entries, 0 to 1940
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   item_name             1941 non-null   object
1   current_price         1941 non-null   float64
2   old_price_element     1941 non-null   float64
3   item_discount_element 1941 non-null   float64
4   item_link             1941 non-null   object
dtypes: float64(3), object(2)
memory usage: 75.9+ KB
```

In [41]:

clean_df.describe()

Out[41]:

	current_price	old_price_element	item_discount_element
count	1941.000000	1941.000000	1941.000000
mean	4359.963936	7419.277950	40.144256
std	5423.488842	9254.119249	8.378224
min	105.000000	154.000000	4.000000
25%	1220.000000	2050.000000	38.000000
50%	3078.000000	5032.000000	41.000000
75%	4884.000000	8193.000000	45.000000
max	54685.000000	79920.000000	54.000000

In [42]:

clean_df.sort_values(by = 'current_price',ascending=False).head(10)

Out[42]:

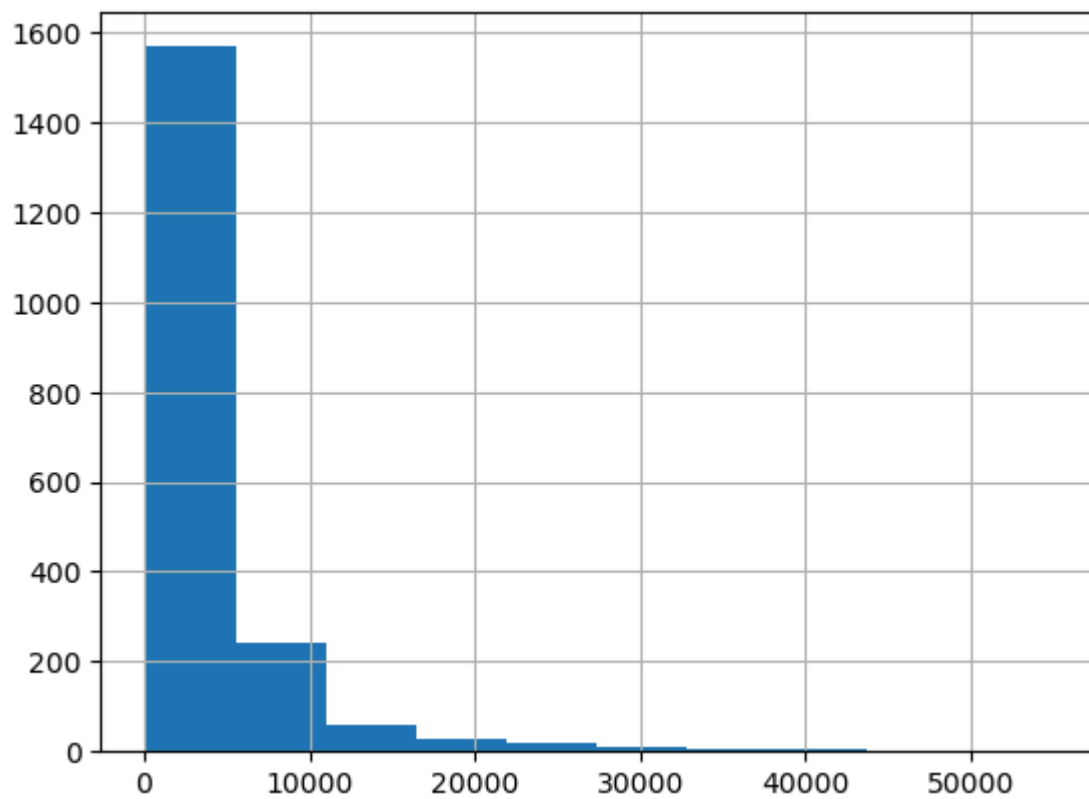
	item_name	current_price	old_price_element	item_discount_element	iter
969	Automatic Flushing Massage Intelligent Toilet ...	54685.0	78122.0	30.0	/generic-auto flushing-ma: intel
1867	U Shape Automatic Cleaning Heating Toilet Clea...	50061.0	71517.0	30.0	/generic-u-s automatic-cle: heatin
1908	Oil Spring Rear Shock Absorber 210mm	47286.0	67552.0	30.0	/generic-oil-s rear-s absorber-210
75	Laptop Parts For MSI GP62 6QG GV62 GP62MVR GL6...	46398.0	79920.0	42.0	/generic-la parts-for-msi- 6qg-gv6
1587	4 8GB 4GB 16GB 32GB 2133mhz 2400mhz 2666mhz 32...	44606.0	76812.0	42.0	/generic-4-8gb 16gb- 2133mhz-240
1654	Modern LED Panda Lamp Kids Room Balloon Ceilin...	42432.0	73112.0	42.0	/generic-mc led-panda- kids-room-
1693	ESP-WROOM-32 ESP32/ESP32S/ESP32U Module 150 Mb...	41825.0	72076.0	42.0	/generi wroo esp32esp32ses mc
750	2x NT75 3/5pin Gasket Mechanical Keyboard DIY ...	39410.0	78819.0	50.0	/generic-2x 35pin-g mechanical-k
1231	2x NT75 3/5pin Gasket Mechanical Keyboard DIY ...	39410.0	78819.0	50.0	/generic-2x 35pin-g mechanical-k
1500	UII Thin Wood Led Ceiling Lamp 15W 48W Modern ...	34987.0	60236.0	42.0	/generic-ul wood-led-c lamp-15

In [43]: `clean_df.sort_values(by = 'item_discount_element',ascending=False).head(10)`

Out[43]:

	item_name	current_price	old_price_element	item_discount_element	item_link
5	Fashion Men Shoes Loafers Casual Shoes Male Sh...	799.0	1743.0	54.0	/fashion-men-shoes-loafers-casual-shoes-male-s...
1930	Fashion Boys Shoes Girls Shoes Children's Canv...	579.5	1200.0	53.0	/fashion-boys-shoes-girls-shoes-childrens-canv...
1872	3Pcs/Set Sister Necklace Set Heart Shape Famil...	225.0	457.0	51.0	/generic-3pcsset-sister-necklace-set-heart-sha...
1008	Feng Shui Black Obsidian Wealth Bracelet	184.0	377.0	51.0	/generic-feng-shui-black-obsidian-wealth-brace...
875	Glitter spool Christmas tree ribbon bandage ca...	172.0	349.0	51.0	/generic-glitter-spool-christmas-tree-ribbon-b...
1690	Silicone Makeup Brush Cleaner Puff Beauty Egg...	895.0	1790.0	50.0	/generic-silicone-makeup-brush-cleaner-puff-be...
978	4x No Tie Athletic Sport Sneakers Elastic Shoe...	1290.0	2579.0	50.0	/generic-4x-no-tie-athletic-sport-sneakers-ela...
981	8008 Rubber Open Watch Back Cover Bottom Ball ...	700.0	1409.0	50.0	/generic-8008-rubber-open-watch-back-cover-bot...
1563	New SR200 smart ring HR monitoring BO monitori...	10212.0	20424.0	50.0	/generic-new-sr200-smart-ring-hr-monitoring-bo...
980	5x DC Power Plug 5.5mm X 2.1mm Male Barrel Ada...	870.0	1739.0	50.0	/generic-5x-dc-power-plug-5.5mm-x-2.1mm-male-b...

In [44]: `clean_df['current_price'].hist()`Out[44]: `<Axes: >`



```
In [45]: ##exporting the cleaned data to a csv  
clean_df.to_csv("cleaned_data.csv", index = False)
```

```
In [ ]:
```